EIAH

2025 Lille

Environnements Informatiques pour l'Apprentissage Humain

Actes de la douzième conférence sur les Environnements Informatiques pour l'Apprentissage Humain (EIAH)

La conférence pluridisciplinaire francophone sur la conception et l'analyse des environnements numériques pour l'éducation et la formation

Édités par Élise Lavoué, Margarida Romero et Yvan Peter

Du 10 au 13 juin 2025
Université de Lille, Lille
France

La douzième conférence EIAH 2025 a été organisée par l'Université de Lille
sous l'égide de l'ATIEF (Association des Technologies de l'Information pour l'Éducation et la Formation)

Table des matières

1 Comités 7

2 Introduction 11

3 Le mot du comité d'organisation 15

4 Conférenciers invités 17

5 Journée ateliers 19

6 Articles longs 23

L'hybridation de la formation professionnelle. Quelle représentation en ont les formateur 3 ans après la pandémie de la Covid-19 ?
Jordan Ajouaou 24

Analyse de l'impact d'une application de jeu sérieux sur les élèves des lycées d'enseignement agricole
Luis Alberto Pinos Ullauri, Catherine Pons, Azzeddine Benabbou et Julien Broisin 40

La Pertinence du Catalogue du Compte Personnel de Formation, vue au Prisme du Comportement des Utilisateurs
Matthieu Cisel 58

Évaluation de différents critères pour la formation de groupes dans le contexte du pair programming distribué : expérimentation dans l'enseignement supérieur
José Colin, Sébastien Hoarau et Julien Broisin 74

Rendre explicable un moteur de recommandations pédagogiques pour permettre une co-évolution système-enseignant
Chloé Conrad, Rémi Venant, Nathalie Guin et Marie Lefevre 92

Apprentissages autorégulés et expérience dans un contexte de formation hybride universitaire
Guelaybe Enosch Djiezion 110

Des compétences aux savoir-faire : proposition d'extensions au métamodèle de COMPER et à MOODLE
Patrick Etcheverry, Thierry Nodenot, Marta Toribio, Christophe Marquesuzaà, Jennifer Lesbats, José Manuel Cornejo Lupa et Benjamin Fontaine ... 126

Analyse semi-automatique des interactions dans une communauté de pratique : influence du type de rencontre sur la dynamique collective
Sabrin Housni, Audrey Kumps, Camelia Pissens, Karim Boumazguida, Gaetan Temperman et Bruno De Lièvre 140

Conception et évaluation : relever les défis de la recherche collaborative menée sur les jeux éducatifs
Mariem Jaouadi, Estelle Prior, Eric Sanchez et Nadine Mandran 158

Modélisation et décision de rétroactions épistémiques dans l'environnement Algo-Python
Sébastien Jolivet, Badmavasan Kirouchenassamy, Amel Yessad et Vanda Luengo 172
Analyse des profils d'utilisation de ChatGPT chez les étudiants de 12 à 29 ans en Fédération Wallonie-Bruxelles
Audrey Kumps, Melody Di Pietro, Sabrin Housni et Bruno De Lièvre 192
Analyse de l'engagement de l'étudiant lié à la difficulté d'items d'évaluation : Une étude d'EdNet fondée sur l'apprentissage automatique
Mohamed Lamgarraj, Céline Joiron, Aymeric Parant et Gilles Dequen 210
Apprendre l'anatomie en dessinant sur tablette : effets du type d'activité et des capacités visuospatiales sur l'apprentissage et les perceptions d'étudiant·es en kinésithérapie
Laura Leconte et Eric Jamet 226
Quand utilité effective et perception de l'utilité diffèrent : Le cas d'un jeu sérieux pour l'entraînement d'enfants malentendants appareillés
Maelys Le Magadou, Marielle Deriaz, Gregory Valentini, Lou Baussand, Matilde Fritz-Legrand, Florent Glück, Mireille Bétrancourt et Angélica Perez-Fornos 240
Une hybridation repensée : Définition, modèle, méthode
Benoît Martinet, Alexis Lebis, Mathieu Vermeulen, Nadine Mandran et Nour El Mawas 256
Impact des feedbacks immédiats sur les apprentissages en chimie dans un environnement de réalité virtuelle
Valentin Midez, Audrey Serna, Florent Diet, Victor Oliva Torres et Élise Lavoué 274
Enjeux juridiques de l'utilisation d'une forge logicielle pour la création et le partage d'EIAH et de ressources pédagogiques en primaire et secondaire
Ezra Nellis, Thierry Forest, Bertrand Marne, Fanny Dessainjean et Iza Marfisi-Schottman 292
Recommander pour Soutenir l'Exploration Urbaines et le Partage d'Expériences entre Étudiants Internationaux
Madjid Sadallah et Marie Lefevre 308
Intégration responsable de l'IA Générative dans l'Éducation : proposition d'un plan d'actions stratégiques dirigé par les risques liés aux questions éthiques
Rémi Saurel, Franck Silvestre, Jean-Baptiste Raclet et Emmanuel Lescure ... 324

7 Articles courts 343

Tirer parti des hallucinations en réévaluant la stratégie de l'agent pédagogique perturbateur à l'ère de l'IA générative
Emmanuel G. Blanchard, Jean-Christophe Callahan, Amine Kessemtini, Lucas Schmitt, Naif Asswiel et Idir Akretche 344
Etude des effets d'un dispositif de formation hybride sur les compétences à s'orienter des étudiants du supérieur
Madina Bouamama, Rawad Chaker et Stéphane Simonian 354

Scraping d'Environnements Numériques de Travail et Analyse de Curriculums : une Etude de Cas
Matthieu Cisel et Camille Bidaud 366

Évaluation des réponses courtes à l'aide de la similarité de phrases
Michel Desmarais, Ovide Bertrand Kuichua Kamikaze dem et Arman Bakhtiari 376

Simulations scientifiques pour les travaux pratiques : choix de conception et affordances pédagogiques
Cédric D'Ham, Sébastien Rebaudo et Rémi Colin de Verdière 388

Effet de la compétition-compétitivité vs coopération-coopérativité sur la recherche d'informations et l'apprentissage sur Internet
Cheyenne Dosso, Jean Sébastien Vayre, Mohamed Benlamine, Tiffany Morisseau et Christophe Heintz 400

Prédiction du risque d'échec dans un MOOC : une approche basée sur l'analyse des séries temporelles multivariées
Anass El Ayady, Maxime Devanne, Germain Forestier et Nour El Mawas ... 416

Formation aux compétences douces : vers un outil auteur pour la création d'agents conversationnels basés sur des LLM
Yassine El Boudouri, Walter Nuninger, Frédéric Hoogstoel, Luigi Lancieri, Julian Alvarez et Yvan Peter 428

Développer des services de formation s'adaptant aux conditions atmosphériques hyperlocales : Une preuve de concept avec l'environnement FACTS
Imran Khan, Emmanuel G. Blanchard et Sébastien George 440

Tester son programme lors d'activités de conception de jeux vidéo à l'école
Jérémy Le Du, Julian Alvarez et Daniel Schmitt 452

Epistémologie instrumentée en EIAH : retour sur l'outil SEED
Nadine Mandran et Alexis Lebis 464

De l'itératif à la navigation accompagnée : une évolution des méthodes de conduite de la recherche
Nadine Mandran, Alexis Lebis et Mathieu Vermeulen 474

Utiliser un lexique visuel pour mieux accéder à l'activité au sein d'un EIAH
Inès Plessis-Ouzariah et Trystan Geoffre 486

Étude comparative entre réfugiés ukrainiens et apprenants en situation de guerre sur la créativité dans des tâches de robotique pédagogique
Margarida Romero et Oksana Strutynska 498

La perception d'utilité d'outils d'échafaudage numériques pour soutenir la résolution de problèmes : comparaison entre trois contextes d'application en enseignement supérieur
Chantal Tremblay 504

8 Posters et démonstrations 515

Identifier un environnement numérique propice à l'expérience optimale d'apprentissage : prémices d'une revue de littérature
Marjolaine Chagrot, Anthony Cherbonnier, Jérôme Guérin et Jean Heutte ... 516

La place des mécaniques ludiques dans les Feedbacks : une revue systématique de la littérature
Adrien Chassaing-Monjou, Alexis Lebis, Mathieu Vermeulen, Anthony Fleury et Nour El Mawas 524
Vers une méthode de comparaison sémantique des modèles d'alignements pédago-ludiques réalisés dans le contexte des jeux sérieux
Madeleine Dufrasne, Anthony Fleury et Alexis Lebis 536
Alignement Sémantique d'Apprentissages Critiques par Embeddings Contextuels
Wissal Essalah, Nicolas Delestre, Nicolas Malandain et Damien Olivier 544
Etude exploratoire de la pertinence des traces pour l'évaluation d'un projet de programmation
Christopher Marshall-Breton, Maude Pupin, Yvan Peter et Mirabelle Nebut . 554
Quels niveaux d'assistance l'IA peut-elle offrir aux enseignants pour faciliter le processus de création d'activités multimodales en réalité mixte ?
Mohammed Oussama Seddini, Mohamed Ezzaouia et Iza Marfisi-Schottman . 560
Repenser l'intégration : le jeu comme métaphore expérientielle pour l'apprentissage
Hélène Parmentier et Eric Sanchez 570
Développement et validation d'un assistant pédagogique intelligent pour les personnes enseignantes
Bruno Poellhuber, Michel Desmarais, Floriane Ronzon et Marie-Noëlle Fortin 576
L'application MOBILES : favoriser le partage d'expériences urbaines entre étudiants internationaux
Madjid Sadallah et Marie Lefevre 582
Co-conception d'un jeu collaboratif multimodal pour le développement des compétences spatiales chez les élèves du primaire
Alain Sénécail, Mireille Bétrancourt, Gaëlle Molinari, Eric Sanchez et Fatou-Maty Diouf 590
Modèle de Recommandation Stochastique et de Raisonnement à Partir de Cas Appliqué à AI-VT
Daniel Soto Forero, Marie-Laure Betbeder et Julien Henriet 600
Conception itérative d'un jeu dans le domaine de la formation médicale
Lyvia Streich, Warisara Suksamran, Nicolas Szilas et Claire Dupont 608

9 Index des auteurs 616

10 Partenaires 619

Comités

Comité de programme

Présidentes :

Élise Lavoué, LIRIS, iaelyon School of Management, Université Jean Moulin Lyon 3, France
Margarida Romero, LINE, Université Côte d'Azur, France

Membres :

Marie-Helene Abel, HEUDIASYC, Université de Technologie de Compiègne, France
Franck Amadieu, CLLE, Université de Toulouse, France
Francine Athias, ELLIADD-FR EDUC, France
Vincent Barré, LIUM, Le Mans Université, France
François-Xavier Bernard, EDA, Université Paris Cité, France
Vincent Berry, LIRMM, Université de Montpellier, France
Mireille Betrancourt, TECFA, Université de Genève, Suisse
Catherine Bonnat, TECFA, Université de Genève, Suisse
François Bouchet, LIP6, Sorbonne Université, France
Denis Bouhineau, LIG, Université Grenoble Alpes, France
Laetitia Boulc'h, EDA, Université Paris Cité, France
Nabila Bousbia, LMCS, Ecole nationale Supérieure d'Informatique, Algérie
Julien Broisin, IRIT, Université de Toulouse, France
Armelle Brun, LORIA, Université de Lorraine, France
Pierre-André Caron, CIREL, Université de Lille, France
Thibault Carron, LIP6, Université Paris 6, France
Jean-François Céci, Techné, Université de Poitiers, France
Jean-François Cerisier, TECHNE, Université de poitiers, France
Ronan Champagnat, L3i, Université de La Rochelle, France
Chantal Charnet, LHUMAIN, Université Paul-Valéry Montpellier, France
Raphaëlle Crétin-Pirolli, CREN, Le Mans Université, France
Colin De La Higuera, LS2N, Université de Nantes, France
Nicolas Delestre, LITIS, INSA Rouen Normandie, France
Bruno De Lièvre, Université de Mons, Belgique
Christian Depover, Université de Mons, Belgique
Michel Desmarais, École Polytechnique de Montreal, Canada
Cyrille Desmoulins, LIG, Université Grenoble Alpes, France
Christophe Després, LIUM, Le Mans Université, France
Philippe Dessus, LaRAC, Université Grenoble Alpes, France
Hassina El Kechai, TECHNE, Université de poitiers, France
Nour El Mawas, CREM, Université de Lorraine, France
Benoît Encelle, LIRIS, Université Lyon 1, France

Cédric Fluckiger, CIREL, Université de Lille, France
Thierry Geoffre, Haute École Pédagogique Fribourg, Suisse
Sébastien George, LIUM, Le Mans Université, France
Jean-Marie Gilliot, Lab-STICC, IMT Atlantique, France
Isabelle Girault, LIG, Université Grenoble Alpes, France
Monique Grandbastien, LORIA, Universite de Lorraine, France
Dominique Groux-Leclet, MIS, Université de Picardie, France
Brigitte Grugeon-Ally, LDAR, Inspé de Créteil, UPEC
Olivier Grugier, EDA, INSPE de Paris, France
Nathalie Guin, LIRIS, Université de Lyon, France
Ludovic Hamon, LIUM, Le Mans Université, France
Sébastien Iksal, LIUM, Le Mans Université, France
Stéphanie Jean-Daubias, LIRIS, Université de Lyon, France
Celine Joiron, MIS, University of Picardie, France
Sébastien Jolivet, IUFE & TECFA, Université de Genève, Suisse
Jean-Marc Labat, LIG, Université Paris 6, France
Yacine Lafifi, Université 8 May 1945 de Guelma, Algerie
Pierre Laforcade, LIUM, Le Mans Université, France
Sébastien Lallé, LIP6, Sorbonne Université, France
Alexis Lebis, CERI SN, IMT Nord Europe, France
Marie Lefevre, LIRIS, Université Lyon 1, France
Dominique Lenne, Heudiasyc, Université de Technologie de Compiègne, France
Paul Libbrecht, IUBH Fernstudium, Allemagne
Mathieu Loiseau, LIRIS, INSA Lyon, France
Domitile Lourdeaux, Heudiasyc, Université de technologie de Compiègne, France
Vanda Luengo, LIP6, Sorbonne Université, France
Nadine Mandran, LIG, Université Grenoble Alpes, France
Iza Marfisi-Schottman, LIUM, Le Mans Université, France
Bertrand Marne, LIUM, Le Mans Université, France
Pascal Marquet, Université de Strasbourg, France
Agathe Merceron, Berliner Hochschule für Technik, University of Applied Sciences, Allemagne
Christine Michel, Techne, Université de Poitiers, France
Najoua Mohib, LISEC, Université de Strasbourg, France
Gaëlle Molinari, TECFA, Université de Genève, Suisse
Mathieu Muratet, LIP6, Sorbonne Université, France
Roger Nkambou, Université du Québec à Montréal, Canada
Sandra Nogry, Paragraphe, Université Cergy-Pontoise, France
Lahcen Oubahssi, LIUM, Le Mans Université, France
Chrysta Pelissier, LHUMAIN, Université de Montpellier 3,France
Claire Peltier, Université Laval, Canada
Mar Pérez-Sanagustin, IRIT, Université de Toulouse, France
Claudine Piau-Toffolon, LIUM, Le Mans Université, France
Laëtitia Pierrot, CREN, Université de Franche-Comté, France
Franck Poirier, Lab-STICC, Université Bretagne-Sud, France
Issam Rebai, Lab-STICC, IMT Atlantique, France

Christophe Reffay, Elliadd, Université de Franche-Comté, France
Robert Reuter, AES, Université de Luxembourg, Luxembourg
Stéphanie Reyssier, ECP, Université Lumière Lyon 2, France
Normand Roy, Université de Montréal, Canada
Eric Sanchez, TECFA, Université de Genève, Suisse
Yann Secq, CRIStAL, Université de Lille, France
Karim Sehaba, LIRIS, Université Lyon 2, France
Audrey Serna, LIRIS, INSA Lyon, France
Franck Silvestre, IRIT, Université Toulouse 1 Capitole, France
Grégory Smits, Lab-STICC, IMT Atlantique, France
Denise Sutter Widmer, TECFA, Université de Genève et HEP-Vaud, Suisse
Nicolas Szilas, TECFA-FPSE, Université de Genève, Suisse
Gaëtan Temperman, Université de Mons, Belgique
André Tricot, Epsylon, Université Paul Valéry Montpellier 3, France
Lucile Vadcard, LARAC, Université Grenoble Alpes, France
Rémi Venant, LIUM, Le Mans Université, France
Mathieu Vermeulen, CERI SN, IMT Nord Europe, France
Amel Yessad, LIP6, Sorbonne Université, France

Coordinateurs des ateliers :

Matthieu Branthôme, IRISA, Université de Rennes, France
Marielle Léonard, CRIStAL, Université de Lille, France

Comité d'organisation :

Président :

Yvan Peter, CRIStAL, Université de Lille, France

Membres :

Maude Pupin, CRIStAL, Université de Lille, France
Frédéric Hoogstoel, CRIStAL, Université de Lille, France
Mirabelle Nebut, CRIStAL, Université de Lille, France
Yann Secq, CRIStAL, Université de Lille, France
Walter Nuninger, CRIStAL, Université de Lille, France
Julian Alvarez, CRIStAL, Université de Lille, France
Marielle Léonard, CRIStAL, Université de Lille, France
Yassine El Boudouri, CRIStAL, Université de Lille, France
Christopher Marshall-Breton, CRIStAL, Université de Lille, France
Pierre-André Caron, CIREL, Université de Lille, France
Mathieu Vermeulen, CERI SN, IMT Nord Europe, France
Alexis Lebis, CERI SN, IMT Nord Europe, France

Introduction

Élise Lavoué, LIRIS
Université Jean Moulin Lyon 3, iaelyon School of Management
elise.lavoue@univ-lyon3.fr

Margarida Romero, LINE
Université Côte d'Azur
margarida.romero@unice.fr

La conférence de la communauté Environnements Informatiques pour l'Apprentissage Humain (EIAH) 2025 s'est tenue à Lille du 10 au 13 juin, marquant la douzième édition de cet événement et rassemblant plus de 120 participants. Dans la continuité des éditions précédentes, cette rencontre s'inscrit dans une dynamique interdisciplinaire, rassemblant des travaux issus de la didactique, de l'informatique, de la psychologie, des sciences de l'éducation, des sciences du langage, des sciences de l'information et de la communication, de la sociologie, et d'autres champs contribuant à la compréhension et à la transformation des environnements numériques d'apprentissage et de formation.

La thématique de l'édition 2025 a été centrée sur l'expérience d'apprentissage et d'enseignement, dans toute sa richesse et sa complexité. Loin de se réduire à l'atteinte d'objectifs pédagogiques, cette expérience s'inscrit dans des contextes multiples, où interviennent des processus socio-culturels, cognitifs, didactiques, affectifs et métacognitifs. Il s'agit ainsi de considérer, à travers les EIAH, ce qui fait expérience pour les apprenant·es comme pour les enseignant·es : l'engagement, la motivation, les émotions, le bien-être, l'auto-régulation, la co-régulation, mais aussi les formes de soutien, d'orchestration ou de médiation mises en œuvre. Lors des présentations, posters et démonstrations au cours de la conférence, les travaux de la communauté ont questionné les EIAH tant dans leur conception, leur intégration aux différents contextes éducatifs, leurs usages, leurs capacités à tirer bénéfice des avancées de l'intelligence artificielle, mais aussi de l'évaluation tant des dispositifs que des apprentissages. Certains travaux ont également abordé la dimension éthique, explicable et soutenable de ces technologies, à toutes les étapes de leur cycle de vie.

L'hybridation des dispositifs, la diversification des contextes éducatifs, les évolutions professionnelles, les usages mobiles ou immersifs, l'émergence d'agents artificiels ou d'outils d'analyse automatisée sont autant d'enjeux abordés dans les contributions réunies cette année. À travers elles, c'est une attention renouvelée portée aux conditions concrètes de l'expérience vécue, à la fois par les apprenant·es et par les enseignant·es, qui se donne à voir. En mettant en dialogue les dimensions techniques, théoriques, contextuelles et humaines des EIAH, les travaux présentés contribuent à une compréhension plus fine des dynamiques à l'œuvre dans les situations d'apprentissage et de formation. La place prise par l'intelligence artificielle dans les processus d'enseignement et d'apprentissage a été analysée aussi bien dans une perspectives des sciences humaines et sociales (SHS) que des sciences informatiques. Des chercheurs ont présenté des dispositifs intégrant des approches centrées sur l'apprentissage machine mais aussi sur les avancées dans les domaines de l'IA générative, les learning analytics, la modélisation des tâches d'apprentissage, de l'apprenant et de l'analyse de données (semi)automatique pour la recherche.

Au-delà de la diversité et l'interdisciplinarité des objets d'étude présentés lors de la conférence EIAH 2025, ce sont aussi les méthodes, les formes de collaboration entre chercheur·es, enseignant·es, ingénieur·es, apprenant·es, ainsi que les relations entre recherche et pratique qui ont nourri les échanges. L'attention portée à la conception, à l'évaluation et à l'accompagnement des usages réels, dans leurs contextes d'émergence, a permis de susciter des réflexions à la fois ancrées et critiques.

Le programme scientifique de cette édition a mis en valeur la diversité des formats de contributions et l'intensité des échanges. 19 communications d'articles longs et 15 communications correspondant à des articles courts ont permis de partager des résultats de recherche consolidés ou en cours d'élaboration. 12 posters scientifiques et 2 démonstrations ont offert des moments privilégiés pour présenter des dispositifs, des prototypes et des retours d'expérience dans un cadre interactif et convivial entre les participants. Le processus de sélection des articles s'est déroulé dans la continuité de l'édition précédente, avec différentes étapes impliquant 95 chercheurs, équilibrés STIC-SHS (dont 18 membres seniors) :

— Affectation d'un méta-relecteur par article de la discipline de l'article (membre senior)
— Évaluation par 3 relecteurs : deux du domaine des auteurs et un d'un autre domaine.
— Rebuttal : réponse des auteurs aux relectures
— Échanges entre les relecteurs et le méta-relecteur pour proposer une décision finale, accompagnée de recommandations pour des révisions par les auteurs : acceptation/refus/reclassement

Nous présentons en tableau 1 le détail des soumissions proposées et acceptées pour chaque format. En complément, la figure 1 illustre ce résultat pour chacune des thématiques renseignées par les auteurs lors de la soumission (pour celles ayant au moins 4 soumissions correspondantes). Cette édition de la conférence est marquée par un très grand nombre de soumissions (70), illustrant le dynamisme de la communauté de recherche francophone dans le domaine des EIAH. Cette édition a été l'occasion de modifier la longueur imposée pour chaque format, limitant en nombre de mots et non plus en nombre de pages. Cette décision a été motivée par la volonté de permettre aux auteurs d'intégrer les tableaux, figures et références illustrant et positionnant de manière pertinente leurs travaux. Ces actes intègrent ces différents travaux en suivant l'ordre alphabétique du nom de famille de leur premier auteur.

Tableau 1 – Soumissions acceptées, reclassées, rejetées ou retirées par les auteurs.

Type	Soumis	Acceptés	Acceptés en courts	Acceptés en posters	Rejetés	Retirés	Publiés
Long	43	20	7	3	13	1	19
Court	21	8		7	6		15
Poster	5	2			3		12
Démo	2	2					2
Total	70	31	7	10	22	1	48

Les conférences invitées de Jean-Marie Burkhardt et d'Engin Bumbacher ont été des temps forts de la conférence, offrant des perspectives complémentaires sur la conception des environnements numériques pour l'apprentissage. Jean-Marie Burkhardt, directeur de recherche à l'Université Gustave Eiffel, a partagé un regard rétrospectif sur plusieurs décennies de recherches en psychologie ergonomique appliquées à la réalité virtuelle et augmentée dans le cadre de l'apprentissage professionnel, mettant en lumière les apports, les limites et les défis actuels de ces technologies immersives. De son côté, Engin Bumbacher, professeur à la Haute École Pédagogique du canton de Vaud, a proposé une réflexion articulée entre théorie et pratique, en interrogeant la conception des outils pédagogiques comme médiation possible entre les cadres théoriques de l'apprentissage et les réalités professionnelles des enseignant·es.

Ces temps scientifiques ont été renforcés par des moments conviviaux forts dès le cocktail de bienvenue du premier jour, et le dîner de gala organisé au couvent des Minimes, un lieu privilégié,

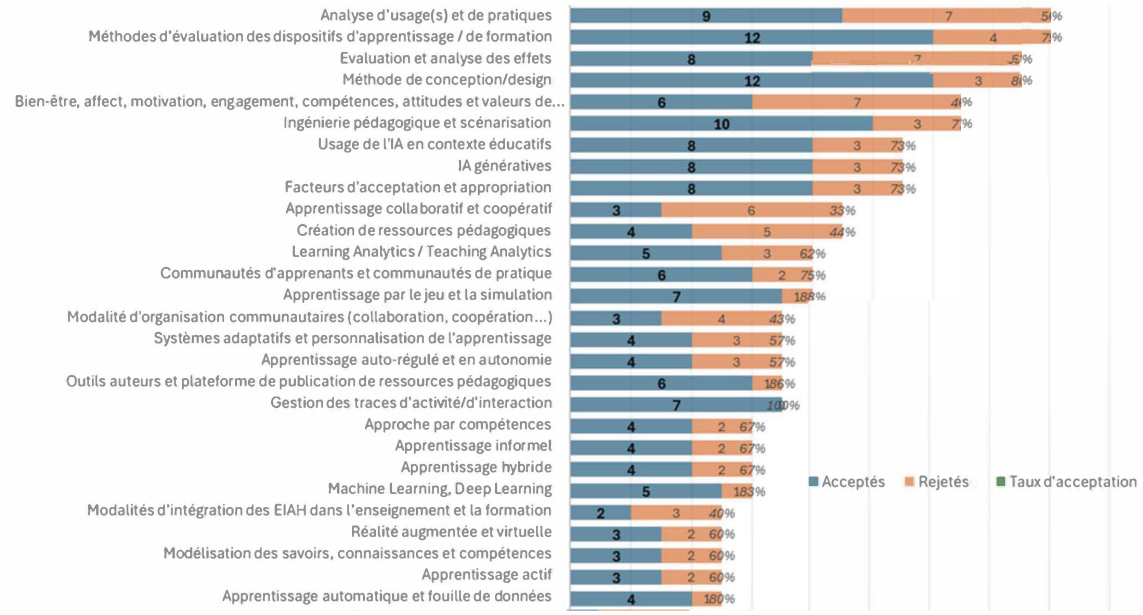


Fig. 1 – Thématiques classées par nombre de soumissions.

qui a permis aux participant·es de se retrouver dans un cadre convivial, facilitant les échanges formels et plus informels qui font la vie d'une communauté scientifique.

Nous tenons à remercier le comité d'organisation, et tout particulièrement Yvan Peter, pour l'excellente collaboration au cours de la préparation de la conférence, les membres du comité de programme pour leur implication dans la garantie d'une conférence de grande qualité scientifique, et chacun·e des participants de la conférence EIAH 2025 pour leurs contributions aux avancées scientifiques et à la vitalité de la communauté EIAH. Nous remercions également l'ATIEF qui nous a fait confiance dans l'organisation de la conférence et accompagné dans les différentes étapes du processus de sélection des publications. Nous tenons enfin à remercier le jury du Prix Monique Grandbastien pour son investissement tout au long de la conférence.

Le mot du comité d'organisation

Yvan Peter, CRIStAL
Université de Lille
yvan.peter@univ-lille.fr

La dernière fois que la conférence EIAH s'est déroulée dans nos contrées, c'était en 2011 à l'Université de Mons. C'est donc avec un immense plaisir que nous accueillons la communauté EIAH à l'Université de Lille en 2025.

Lorsque l'ATIEF m'a sollicité pour l'organisation de cet événement, j'avais conditionné mon acceptation à l'engagement de l'équipe NOCE dans cette aventure, et il n'y a eu aucune hésitation de leur part. Nous avons tout naturellement sollicité le CIREL, et il n'a pas fallu longtemps pour que quelqu'un propose d'impliquer nos collègues de l'IMT Nord Europe. Cette collaboration nous a permis de constituer un comité d'organisation conséquent qui démontre que la communauté EIAH est bien représentée dans le Nord, tant du côté informatique que du côté des sciences de l'éducation et de la formation.

La participation a été remarquable. Pour la journée ateliers nous avions 126 participant·es réparti·es parmi 10 ateliers couvrant des thématiques très diverses. Pour la conférence principale, nous comptions 120 personnes inscrites. Au total, cela nous fait 167 personnes inscrites à l'ensemble de l'événement !

Nous remercions chaleureusement l'ATIEF pour sa confiance, ainsi que les deux présidentes du comité de programme, Élise Lavoué et Margarida Roméro, avec lesquelles la collaboration nous a permis d'avancer sereinement dans l'organisation. Nous remercions également Marielle Léonard et Matthieu Branthôme qui ont géré la journée des ateliers.

Cette conférence bénéficie du soutien institutionnel de la Région Hauts-de-France et de l'Université de Lille. L'IMT Nord Europe, Polytech'Lille et les laboratoires CIREL et CRIStAL ont également contribué, que ce soit financièrement ou par un accompagnement administratif et logistique. Le soutien de ces différents acteurs est essentiel à la réussite de l'organisation, et nous les remercions chaleureusement.

J'espère que les participant·es ont profité pleinement de riches échanges scientifiques (ou non...) et qu'ils et elles ont apprécié l'accueil que nous leur avons réservé.

Conférenciers invités

La conception des outils pédagogiques comme médiateur entre théories de l'apprentissage et enseignant·e·s

Engin Bumbacher, Professor in CS Education and Education Technology
Haute école pédagogique Vaud
engin.bumbacher@hepl.ch

Résumé

Les outils technopédagogiques jouent un rôle important dans l'enseignement — à la fois par ce qu'ils rendent possible, et par les conceptions de l'apprentissage qu'ils véhiculent.

Dans un premier temps, je montre comment chaque décision de conception a un impact direct sur les processus d'apprentissage et d'enseignement, en reflétant toujours une certaine vision de l'apprentissage, implicite ou explicite. Les théories de l'apprentissage sont donc essentielles pour guider la conception, et je montre comment elles peuvent être intégrées concrètement pour produire des outils qui alignent les expériences d'apprentissage avec les intentions pédagogiques, en créant des expériences d'apprentissage riches et exigeantes.

Dans un second temps, j'analyse comment la conception technopédagogique transforme la manière dont les enseignant·e·s pensent l'apprentissage — que ce soit par l'usage d'outils issus d'une démarche de design guidée par la théorie, ou en s'impliquant activement en tant que concepteurs. Je montre comment engager les enseignant·e·s dans un processus de conception permet de révéler les tensions entre intentions, expériences pratiques et conceptions de l'apprentissage, et comment ce processus peut servir de levier pour approfondir et transformer leur compréhension de l'apprentissage et de l'enseignement.

À partir d'exemples concrets, cette conférence propose une vision de la conception technopédagogique comme interface vivante entre savoirs théoriques et savoirs professionnels, capable de transformer ce que l'on conçoit, ce que l'on enseigne, et ce que l'on comprend de l'apprentissage.

Biographie



Engin Bumbacher est professeur en didactique de l'informatique et en technologies éducatives à la HEP Vaud (Suisse). Il mène un programme de recherche appliquée axé sur la conception de technologies et de services éducatifs innovants, afin de soutenir les enseignants dans le développement de pratiques pédagogiques engageantes, équitables et fondées sur les compétences en STEM.

Ses travaux portent sur trois axes : (1) le développement d'outils pour l'enseignement et l'évaluation de la résolution de problèmes informatiques, entre autre dans le cadre de l'évaluation PISA 2025 de l'OCDE ; (2) l'intégration de projets STEM du monde professionnel dans les cours

de sciences au primaire et au secondaire à travers des nouveaux dispositifs pour rendre visible des possibles futures aux élèves ; (3) la création de nouveaux formats de formation ancrés dans les fablabs pour accompagner les enseignant·e·s dans la conception de dispositifs pédagogiques, entre autre un projet en partenariat avec le Teachers College Columbia University. Il est titulaire de diplômes de l'ETH Zurich (physique, neuroscience) et d'un doctorat en "Learning Sciences and Technology Design » de l'Université Stanford.

Psychologie ergonomique de la réalité virtuelle et augmentée pour l'apprentissage professionnel : un point de vue sur quelques décennies de recherches et d'applications

Jean-Marie Burkhardt, Directeur de recherche
LaPEA, Université Gustave Eiffel, IFSTTAR
Université de Paris
jean-marie.burkhardt@univ-eiffel.fr

Résumé

Dans cette présentation, je retracerai les travaux menés dans le champ de la conception et de l'évaluation des environnements virtuels et des dispositifs de réalité augmentée pour l'apprentissage, avec un focus sur les contenus professionnels. En m'appuyant sur des exemples des recherches et des dispositifs développés sur une période de près de 30 ans, je tâcherai d'analyser les évolutions en termes de questionnement de recherche, d'applications et d'approches, et proposerai d'en tirer quelques implications pour nourrir la prospective sur ces technologies en cours d'émergence.

Biographie



Jean-Marie Burkhardt est directeur de recherche à l'Université Eiffel au sein du Laboratoire de Psychologie et d'Ergonomie Appliquées. Il développe des recherches selon deux principaux axes thématiques : (1) l'évaluation et le développement des technologies émergentes, en particulier dans les domaines de la réalité virtuelle et de la réalité augmentée ; (2) la psychologie des activités de conception et des concepteurs, avec un intérêt particulier pour la dimension collaborative de ces activités et leur outillage. Il a par ailleurs présidé le groupe de travail de l'ANSES sur les effets sanitaires de l'exposition aux technologies de réalité virtuelle et/ou augmentée.

Journée ateliers

Matthieu Branthôme, IRISA
Université de Rennes
matthieu.branthome@univ-rennes.fr

Marielle Leonard, CRIStAL
Université de Lille
marielle.leonard@univ-lille.fr

La journée des ateliers de la conférence EIAH 2025, qui s'est tenue le 15 juin à Lille, a offert un espace de réflexion collective et de dialogue scientifique autour de thématiques clés pour notre communauté. Conçus comme des lieux d'échanges approfondis, ces ateliers ont permis de favoriser la confrontation de perspectives pluridisciplinaires, d'expériences de terrain et de travaux de recherche.

Les 10 propositions retenues ont pris des formats variés, sous forme de symposiums ou de workshops. Ils visaient à stimuler la participation active des intervenant·es et des participant·es, et à encourager l'exploration critique des problématiques discutées. Les thématiques abordées lors de cette journée s'organisaient autour de quatre grands axes : la conception et l'évaluation des EIAH, l'apprentissage du numérique et les facteurs humains impactant les processus d'apprentissage, les démarches pluridisciplinaires et inclusives, et enfin, les usages et enjeux liés à l'intelligence artificielle générative dans le champ des EIAH.

Cette diversité témoigne de la vitalité de la recherche en EIAH et de sa capacité à articuler des dimensions technologiques, pédagogiques, méthodologiques et sociétales. Nous exprimons notre reconnaissance à l'ensemble des porteur·euses d'ateliers pour la qualité scientifique des propositions, ainsi qu'aux participant·es pour leur contribution active aux discussions et aux travaux de cette journée. Nous remercions sincèrement le Conseil d'Administration de l'ATIEF pour la confiance qu'il nous a témoignée en nous confiant l'organisation de cette journée d'ateliers.

Conception et évaluation des EIAH

Évaluation et reconnaissance ouverte dans les EIAH

Le symposium *Évaluation et reconnaissance ouverte dans les EIAH* a été porté par Pierre-André Caron, Bruno Delièvre et Maude Pupin. Il a été composé de deux ateliers participatifs, l'un sur la co-création d'un dispositif d'évaluation à la première personne, l'autre sur le processus de création, de diffusion et d'usage des Open Badges.

Visualiser des indicateurs de compétences : objectifs et verrous

Le symposium *Visualiser les indicateurs de compétences : objectifs et verrous* a été organisé par Jean-Marie Gilliot et Matthieu Vermeulen, pour les GT « compétences et EIAH » et « tableaux de bord d'apprentissage » de l'ATIEF. Il a consisté à croiser les préoccupations des experts en approche par compétences et des spécialistes des tableaux de bord d'apprentissage afin d'identifier les verrous qui freinent l'efficacité de ces outils et de proposer des pistes pour une meilleure intégration des indicateurs relatifs aux compétences dans les tableaux de bord d'apprentissage.

GDR IHM – GT Apprentissage et Éducation

L'atelier *GDR IHM*, organisé par Iza Marfisi, Arnaud Prouzeau et Audrey Serna, s'est attaché à s'inscrire dans les thématiques définies dans le GT Apprentissage et Éducation et portée par la communauté IHM et EIAH. Plusieurs environnements numériques ont été explorés et analysés, et une étude méthodologique au sujet des questionnaires qualitatifs a été présentée.

Apprentissage du numérique et facteurs humains impactant l'apprentissage

Apprentissage de la pensée informatique de la maternelle à l'université

Le workshop *APIMU*, organisé par Christophe Declercq, Yann Secq, Julien Broisin et Yannick Parmentier, a abordé l'usage d'instruments pour l'apprentissage de la pensée informatique pour faire obstacle aux stéréotypes de genre, notamment les machines notionnelles, les scénarios UML, les jeux sérieux et les environnements de programmation. Une typologie des erreurs en programmation et leur détection automatique dans un code python a également été présentée.

Soutenir la motivation et l'engagement des apprenants dans les EIAH

L'atelier *Soutenir la motivation et l'engagement des apprenants dans les EIAH* a abordé la question de la prise en compte de la motivation et de l'engagement des apprenants, à travers tout le cycle de vie d'un EIAH, de sa conception à ses usages. Cet atelier a été organisé par Élise Lavoué et Stéphanie Reyssier.

Démarches pluridisciplinaires et inclusives

Environnements informatiques pour l'apprentissage et inégalités numériques

Les articles présentés lors de l'atelier *Environnements informatiques pour l'apprentissage et inégalités numériques* ont abordé l'accessibilité des outils numériques dans les structures éducatives, dans l'objectif d'identifier et promouvoir des outils favorisant une éducation plus inclusive. Cet atelier a été organisé par Sébastien Iksal, Yvan Peter, Florence Bara, Alessandro Bergamaschi en lien avec le RT Éducation.

EIAH et didactique(s) : quelles collaborations et apports mutuels ?

Le symposium *EIAH et didactique(s) : quelles collaborations et apports mutuels ?* a été organisé par Nathalie Guin, Sébastien Jolivet. Plusieurs présentations se sont succédées, qui ont montré les ressorts et les défis de recherches pluridisciplinaires impliquant notamment didacticiens et informaticiens.

Les universités virtuelles : un outil de diffusion du savoir par la formation à distance

Le symposium *Les universités virtuelles : un outil de diffusion du savoir par la formation à distance* a été organisé par Guy Mbatchou, Philippe Pernelle. Les présentations ont soulevé les problématiques de la formation à distance, via des universités virtuelles, notamment en Afrique. Parmi les questions abordées figurent l'hybridation, le maintien de la motivation des apprenants et l'évaluation des compétences.

Problématiques et usages autour de l'intelligence artificielle générative

Ludopédagogie et Intelligence artificielle : innovation, limites et perspectives

L'atelier *Ludopédagogie et Intelligence Artificielle : innovations, limites et perspectives* a été porté par Julian Alvarez et Yvan Peter. L'atelier a été constitué de présentations qui ont abordé l'utilisation de l'intelligence artificielle générative pour la conception de jeux sérieux, parfois combinée avec d'autres technologies innovantes comme la réalité virtuelle.

Des systèmes d'intelligence artificielle pour l'enseignement et l'apprentissage ancrés dans le paradigme de l'intelligence hybride

Le symposium sur les *systèmes d'intelligence artificielle pour l'enseignement et l'apprentissage* ancrés dans le paradigme de l'intelligence hybride a exploré comment l'intelligence hybride peut être intégrée à la conception d'EIAH en utilisant des systèmes d'intelligence artificielle générative. Ce symposium, porté par le consortium international LAVIA, a été organisé par Jean Heutte, Gaëlle Molinari, Michel Desmarais et Bruno Poelhuber.

EIAH
2025 Lille
Environnements Informatiques pour l'Apprentissage Humain

Articles longs

L'hybridation de la formation professionnelle

Quelle représentation en ont les formateurs 3 ans après la pandémie de la Covid-19 ?

Jordan, AJOUAOU[1]

[1] Université de Lille
jordan.ajouaou@gmail.com

Résumé. Cet article porte sur la représentation sociale des cadres de santé formateurs vis-à-vis de l'hybridation de la formation professionnelle. La recherche à l'origine de cet article s'appuie sur une approche psycho-sociale. Définissant la spécificité de l'hybridité d'un dispositif de formation professionnelle ainsi que la structuration d'une représentation sociale, cette étude s'intéresse tout particulièrement à identifier et cartographier les éléments constitutifs de la représentation sociale qu'ont les cadres de santé formateurs d'un dispositif de formation, dans le contexte d'universitarisation de la formation professionnelle ainsi que de la pandémie de la covid-19.

Mots-clés : TICE, Hybridité, Représentation sociale, Formation professionnelle, Analyse prototypique

Abstract. This article looks at the social representation of head nurse teachers with regard to the hybridisation of professional training. The research behind this article is based on a psycho-social approach. Defining the specificity of the hybridity of a professional training system and the structuring of a social representation, this study is particularly interested in identifying and mapping the constituent elements of the social representation of a training system by head nurse teachers, in the context of the universitarisation of professional training and the covid-19 pandemic.

Keywords: ICTE, Hybridity, Social representation, Vocational training, Prototypical analysis

1 Introduction

Au sein des instituts de formation en soins infirmiers (IFSI), évoluent des cadres de santé, soignants de formation. Ces derniers ont pour mission de former les futures générations d'infirmiers à la pratique de leur métier et au développement des compétences nécessaires à l'obtention du Diplôme d'État infirmier. L'hybridation des dispositifs de formation, débutée par l'universitarisation de la formation professionnelle en 2009, et

accélérée par la pandémie de la Covid 19, vient questionner les pratiques d'enseignement. Ce travail cherche à s'intéresser aux représentations des cadres de santé formateurs quant à l'hybridation de la formation infirmière. Dans un premier temps, nous aborderons les concepts d'hybridité et de représentation sociale, puis nous décrirons l'étude menée auprès de 20 cadres de santé formateurs au sein des IFSI de l'Assistance Publique des Hôpitaux de Paris (AP-HP) afin d'identifier et de cartographier les éléments constitutifs du noyau central de leur représentation sociale de l'hybridité de la formation professionnelle.

2 L'hybridité, une complexité à définir

L'étymologie du terme « hybride », trouve son origine dans le terme latin *hybrida* qui signifie mêler (Halpern, 2020). Ce terme indique quelque chose de l'ordre de la complexité. Cette complexité est, en quelque sorte, une combinatoire d'éléments agencés spécifiquement, convoquant, une pensée dialogique, nous permettant « de maintenir la dualité au sein de l'unité » (Morin, 1990). Ainsi, l'hybridité est une association de divers éléments étant en eux-mêmes déjà un « Tout » dans leur singularité, mais constitutifs d'un nouveau « Tout » une fois assemblés. Il y a donc dans ce concept d'hybridité une troisième voie, une place à la complexité, à l'altérité. Ainsi, la pensée binaire n'opère plus, et cet « Autre » issu de l'hybridation, est un produit résultant d'une association-juxtaposition d'éléments, lui conférant son unicité. L'hybridité a alors un caractère unique, rendant les dispositifs de formation hybrides spécifiques et uniques, un tiers-lieu où le temps et l'espace, le réel et le virtuel et la présence et la distance coexistent.

2.1 Les spécialités et temporalités hybrides

Avant de s'intéresser à la notion de tiers-lieu, définissons la dimension spatiale qu'induit ce concept. En effet, l'espace peut être décrit comme deux entités distinctes. Tout d'abord, un espace absolu qui est un « milieu homogène et vide, infini et infiniment divisible, se prêtant indifféremment à n'importe quel mode de décomposition » (Bergson, 1922). Cet espace est par nature une séparation entre des éléments. Par ailleurs, il est possible d'observer un espace dit relatif, beaucoup plus en lien avec son contenu. Ainsi, « l'espace n'est rien d'autre que l'ordre de l'existence simultanée des possibles » (Leibnitz, 1704/1962). L'espace est donc une entité propice à de l'existant ouvrant ainsi le champ des possibles. En mathématiques et en géographie, l'espace est davantage lié à la notion de distance. Cette notion de distanciation, établie une séparation entre soi et les autres, permettant de distinguer des objets coexistant indépendamment de leurs qualités intrinsèques. Ainsi la dimension spatiale d'un dispositif de formation peut être vue davantage comme un espace relatif permettant d'abriter un dispositif de formation rendant ainsi propice l'acte même de former. La formation infirmière étant un dispositif de formation par alternance, il est possible de la concevoir comme une dualité d'espace de formation. Les apprenants sont amenés à changer d'espace et de lieu pour se former. Ainsi, se questionner sur la notion spatiale d'un dispositif de

formation permet d'aborder la complexité de la notion de distance. En effet, il est possible de distinguer six classes de distance. Il y a « les distances spatiale, temporelle, technologique, socioculturelle, socio-économique et pédagogique » (Jacquinot, 1993). Outre la dimension spatiale et temporelle, il est possible d'observer la distance technologique qui est « liée aux facilités ou aux obstacles pour utiliser des artefacts technologiques mis à disposition » (Jézégou, 2022). De même, la distance socioculturelle, est quant à elle, « la séparation entre l'univers de la formation instituée et les exclus du système éducatif » (Jézégou, 2022). La distance socio-économique relève davantage de « l'écart existant entre l'investissement réalisé dans la formation […] et le retour réalisé sur cet investissement » (Jézégou, 2022). Pour finir, la distance pédagogique est liée au différentiel entre deux logiques « celle du sujet en formation […] et celle poursuivie par l'institution éducative » (Jézégou, 2022). Ainsi, l'élément spatial est bien plus vaste qu'il n'y parait. Entrer dans le domaine de l'hybridité, met en lumière le caractère complexe de sa forme. L'espace de formation n'est alors pas uniquement une alternance de modalité spatiale binaire (présence/distance) mais peut plutôt être compris comme un agencement de ces multitudes de distances à l'œuvre au sein d'un dispositif de formation.

La notion temporelle est également une notion relativement complexe. Outre le temps physique intéresserons-nous ici à la subjectivité de cette notion. En effet, « pas de temps objectif sans un sujet pour le percevoir » (Ricœur, 1984). C'est le temps vécu par l'individu qui le rend observable. L'apprenant inscrit dans un processus de formation est confronté au temps présent. Cette conception de la dominance du présent renvoie à une « grandeur infinie, donnée, unique précédant a priori toutes ses pièces » de sorte que « tous les temps perçus soient des parties du temps ». L'imagination transcendantale qui relie sensibilité et entendement est l'instance qui « figure le présent, reconfigure le passé et préfigure le futur » (Kant, 1783/1964, cité dans Dubar, 2008). Ainsi le temps tel qu'il est possible de le percevoir est une analyse au présent des réalités du présent mais aussi celles du passé et celles à venir. En d'autres termes, l'apprenant et le formateur, évoluant dans un dispositif de formation, sont en quelque sorte en permanence dans le présent étant donné que c'est de leur perception présente qu'ils analyseront le temps vécu. Le temps est alors une « durée comme flux continu » (Bergson, 1922). Autrement dit, le temps n'est pas uniquement une mesure temporelle entre un début et une fin mais bien une continuité. Le formateur perçoit le temps en fonction des différentes activités éducatives qu'il est en train de réaliser. De plus, par l'introduction du numérique, « des pratiques sociales hybrides remettent en cause des oppositions bien établies entre les médias et la communication interpersonnelle […], entre les loisirs et le travail, entre l'école et l'auto-apprentissage » (Flichy, 2021). Il est donc possible d'observer un certain éclatement temporel. De plus, la notion de synchronicité et d'asynchronie confère au tiers lieu toute sa dimension hybride. Le présent est donc la temporalité dominante ici. Le formateur peut alors enregistrer les cours, et les mettre à disposition des apprenants. Le formateur réalisant son cours, vit alors son action d'enseignement au présent. Or, le cours mis à disposition par le formateur peut être considéré comme faisant partie du passé, dans le sens où ce dernier est déjà réalisé. L'apprenant, visionnant/écoutant/lisant le cours mis à sa disposition, vit alors son action d'apprentissage lui-même au présent. Ainsi les deux temporalités se voient conjointement adossées du fait de l'asynchronicité du dispositif de formation.

Autrement dit, « la capacité à articuler les formes de temporalités à l'œuvre dans les métiers (apprenant ou formateur) serait donc une capacité dépassant les compétences organisationnelles et impliquant de façon plus large le rapport à soi et le rapport à son environnement » (Varga, 2013). Un lieu où les temporalités et les localisations se voient libérées des dualités traditionnelles : présence/distance et passé/présent, s'apparentant à une « hétérotopie [qui a] le pouvoir de juxtaposer en un seul lieu réel plusieurs espaces, plusieurs emplacements qui sont en eux-mêmes incompatibles » (Foucault, 1967). Cette notion vient définir le tiers lieu hybride dans le sens où ce dernier renferme en son sein toutes les potentialités temporelles et spatiales.

2.2 Le réel, le virtuel et la présence hybride

Au sein d'un dispositif de formation instrumenté par l'outil numérique les formateurs font usage des artefacts mis à leur disposition, plongés dans une certaine virtualité qui « se construit comme une interaction, une négociation entre technologie et utilisateurs » (Proulx, 1994). Ce « dualisme numérique » est la croyance selon laquelle l'espace numérique et le monde physique sont séparés (Jurgenson, 2012 ; Vial, 2016). Cependant, comme le souligne Jézégou (2022) cette opposition réel/virtuel semble non opérante étant donné le contexte hybridé des dispositifs de formation. En effet, le mot virtuel provient du latin *virtualis* signifiant : force ou potentiel. Ainsi, « la *virtus* est à la fois la cause initiale en vertu de laquelle l'effet existe mais aussi ce par quoi la cause continue d'exister, présente virtuellement dans l'effet. Le virtuel n'est donc ni irréel ni potentiel : le virtuel est de l'ordre du réel » (Quéau,1993 ; Jézégou, 2019). La séance pédagogique par visioconférence, le cours virtuel réalisé via l'outil numérique est, dès lors, une expression du réel et non une non-classe de cours. En effet, bien que le regroupement des apprenants et formateur se fasse via un artefact numérique dématérialisant la salle de cours, le contenu du cours lui reste réel, et les échanges également. Ainsi donc, « le virtuel a vocation à s'hybrider au réel, à constituer une sorte de complexe réel-virtuel, une nouvelle réalité composite [...] le virtuel permet d'accoucher du réel » (Quéau,1993). Le réel et le virtuel s'hybrident donc l'un à l'autre au travers d'interfaces, pour, *in fine*, produire du réel, créer de la présence même à distance.

Ces deux notions, dans le contexte d'hybridation d'un dispositif de formation, sont de fait, associées. En effet, le contexte d'hybridation nous fait alors changer de « paradigme passant de la notion de distance qui sépare à celle de proximité qui relie » (Paquelin, 2011). La distance physique n'est plus un obstacle à la proximité. Dans cette dialogique distance/proximité (Jézégou, 2022), la présence s'avère être une notion porteuse d'hybridation. L'introduction des TICE est venue complexifier le concept même de présence. En effet, « le cercle des échanges entre tuteurs et tutorés ne se limite plus à l'espace local mais s'étend bien au-delà des frontières de la géographie universitaire […] depuis l'arrivée des TICE dans l'enseignement supérieur » (Baudrit, 2018). De plus, nous l'avons vu plus tôt, c'est par l'écran qu'un certain rapprochement s'opère. L'écran diffusant une image qui semble être « le contraire de la présence […] L'image télé virtuelle présente le signe d'une « présence » réelle. Quelqu'un de vivant est bien « là », certes virtuellement par l'apparence mais bien réellement par l'attention disponible, la capacité de communiquer et d'agir » (Quéau, 1993). Ainsi, le concept

même de présence revêt un caractère hybride. La présence ne peut donc plus être considérée comme un état (Licoppe, 2012). En effet, la présence physique semble alors ne pas être l'unique modalité pour assurer une présence. Ainsi, deux modalités de la présence spatiale peuvent être observées, l'une physique et l'autre numérique (Paquelin, 2011). Ces éléments amènent à penser qu'il existe « des distances possibles dans la proximité physique et des proximités possibles dans la distance géographique » (Jézégou, 2019). Il apparaît alors que la présence est issue d'un sentiment de présence ressenti par l'apprenant. La présence est alors liée à l'expérience subjective de l'apprenant, au cours de la séance animée par le formateur (Witmer et Singer, 1998). Ainsi, le formateur de par son action peut susciter le sentiment de présence auprès de ses apprenants et ce même lors d'une session d'enseignement à distance. En résumé, les notions de distance et de proximité ne sont donc pas en opposition mais bien constitutives d'une présence comme tiers-hybride.

3 Les représentations sociales

3.1 La structure et les fonctions

Les représentations ont été étudiées dès le dix-neuvième siècle. Il était dès lors affirmé que « ce qui nous dirige, ce ne sont pas les quelques idées qui occupent présentement notre attention ; ce sont tous les résidus laissés par notre vie antérieure ; ce sont les habitudes contractées, les préjugés, les tendances qui nous meuvent sans que nous nous en rendions compte, c'est, en un mot, tout ce qui constitue notre caractère moral » (Durkheim, 1898). Les représentations sont donc des éléments constitutifs de l'individu. En effet, chercher à comprendre ce qu'est la représentation, « c'est comprendre comment l'esprit façonne les rapports et les institutions, puisque chaque rapport ou institution comporte des images et des notions qui non seulement les expriment, mais jouent le rôle de contraintes sociales servant à ordonner les gens et à trier les choses » (Kalampalikis, 2019). La représentation joue un rôle significatif quant aux rapports sociaux que les individus entretiennent avec leur environnement et le sens qu'ils lui attribuent. De plus, une « dialectique […] est au cœur de la théorie des représentations sociales et [celle-ci] concerne les rapports entre croyances, individus et groupes sociaux » (Moliner et Guimelli, 2015). C'est au croisement de ces multiples notions qu'il faut évoluer, afin de déceler la nature même de la représentation sociale et de son rôle organisateur de l'action d'un individu. La représentation se définit comme « le produit et le processus d'une activité mentale par laquelle un individu ou un groupe reconstitue le réel auquel il est confronté et lui attribue une signification spécifique. La représentation est donc un ensemble organisé d'opinions, d'attitudes, de croyances et d'informations se référant à un objet ou une situation. Elle est déterminée à la fois par le sujet lui-même (son histoire, son vécu), par le système social et idéologique dans lequel il est inséré, et par la nature des liens que le sujet entretient avec ce système social » (Abric, 2003a). Ainsi, la représentation est tout d'abord issue d'un processus mental, qui permet de se représenter le réel. En effet, l'individu va chercher mentalement à concevoir ce qui l'entoure. En ce sens, la représentation cherche à donner du sens à l'inconnu. Cet « inconnu » se voit attribué un sens par l'individu afin que ce dernier puisse l'intégrer comme « connu » de lui. L'ensemble du processus de

représentation s'articule autour d'un noyau central caractérisé comme un « noyau dur » par Mugny et Carugati (1989) ou bien comme un « principe organisateur » par Doise (1990). Ce noyau est une structure organisant les constituants de la représentation et leur donne sens (Flament, 2003). Ainsi donc, la représentation, en tant que structure fondamentale, se présente comme un système complexe, en perpétuelle évolution autour d'un axe central (le noyau). Ce dernier, par sa position centrale et sa stabilité apparente, permet d'assurer quatre fonctions primordiales (Voir Fig. 1).

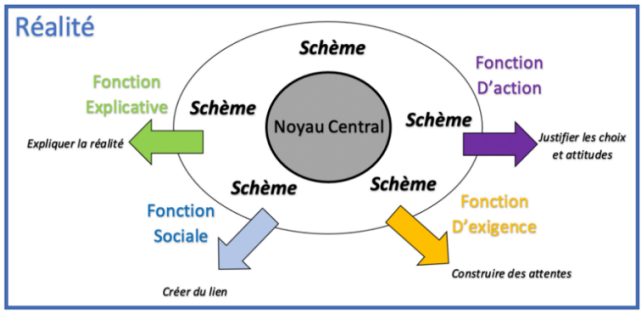


Fig. 1. Structure et fonctions de la représentation (Ajouaou, 2024)

La fonction explicative décrypte et cherche à donner du sens à l'environnement et la fonction d'action cherche à connaitre l'environnement pour pouvoir pleinement agir sur ce dernier. La représentation remplit deux autres fonctions : la fonction sociale (l'individu, évoluant dans une société, produit des interactions sociales) et la fonction d'exigence (représentations propres au groupe, créant une distinction d'attente et d'exigence). Ces fonctions permettent d'appréhender l'environnement dans lequel évolue l'individu au quotidien, mais également de créer des liens sociaux, de motiver les attentes individuelles et de justifier les actions entreprises. La représentation permet à l'individu de comprendre son environnement et d'établir avec celui-ci des liens sociaux par le biais de ses interactions avec les autres. De plus, la représentation donne un cadre de référence induisant une exigence et des attentes spécifiques. Pour finir, la représentation est initiatrice de l'action. Par ses schèmes périphériques la représentation entre en interaction avec la réalité de l'environnement. Par interprétation des éléments inconnus, l'individu s'adapte au réel dans lequel il évolue. Les schèmes ont alors cette double fonction d'interprétation du réel, et de préservation du noyau central, véritables tampons entre la réalité et la représentation que l'individu s'en fait (voir Fig. 2).

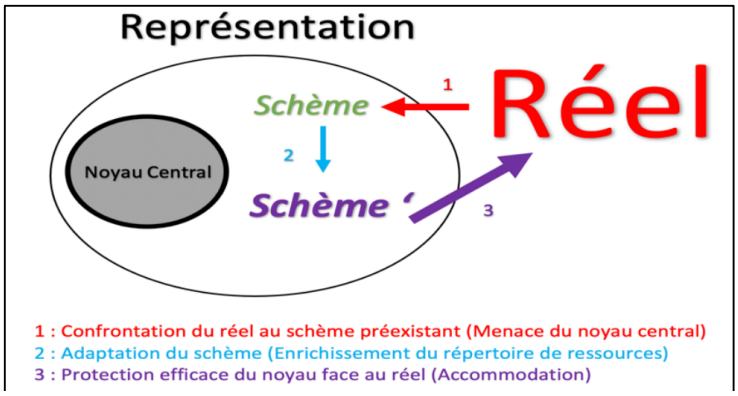


Fig. 2. Double processus d'adaptation/protection du schème (Ajouaou, 2024)

Le schème peut être défini comme « la structure ou l'organisation des actions, telle qu'elle se transfère ou se généralise lors de la répétition de cette action en des circonstances semblables ou analogues » (Piaget et Inhelder, 1966). Ainsi, le schème permet à l'individu d'organiser ses actions par une conceptualisation du réel du fait d'une confrontation répétée. Parallèlement, ils lui donnent également la possibilité de conceptualiser l'action en mobilisant une connaissance prédicative, constituant ainsi un ensemble de structures mentales évolutives et dynamiques. Ainsi, les schèmes, résidant au sein du noyau central, se voient confortablement protégés par ses schèmes périphériques. Réels éléments adaptatifs, les schèmes sont alors des interfaces entre la représentation statique du noyau central, empreinte d'une certaine inertie, et le réel toujours plus mouvant et étrange pour l'individu. Les schèmes peuvent être perçus comme de véritables remparts mouvants, permettant à l'individu de s'adapter à la mutation du réel environnant, mais également, par leur adaptation, de produire de nouvelles connaissances du réel.

3.2 La représentation sociale et le contrôle de son environnement

L'individu, dans une dynamique d'adaptation permanente en périphérie de ses représentations, tente de « pallier les étrangetés, de les réintroduire dans l'espace collectif, en amenant à se rencontrer des visions, des expressions disparates et séparées qui, en un sens, se cherchent […], l'insolite se change en coutumier » (Moscovici, 1976). Son évolution dans le réel incite l'individu, par une adaptation permanente, à appréhender son environnement et les éléments qui le constituent. De plus, « les représentations sociales sont des modalités de pensée pratique orientées vers la communication, la compréhension et la maîtrise de l'environnement social, matériel et idéel » (Jodelet, 1985). C'est-à-dire que l'individu tente, par cette adaptation, non seulement de comprendre son environnement et les éléments qui le constituent, mais il cherche également à le maitriser. Cette prise de contrôle peut s'exprimer par une orientation pratique. L'individu cherche donc à communiquer afin de s'inscrire dans un environnement social. De plus, la maitrise des idées et du matériel à sa disposition lui permet de s'assurer d'un degré significatif de contrôle sur sa réalité. Ainsi, l'individu,

cherche par la maitrise de techniques ou d'outils de pouvoir comprendre, contrôler, voire maitriser son environnement. La représentation est également un vecteur d'appropriation. En effet, l'individu va se confronter au réel, et parfois l'utilisation d'instruments lui sera nécessaire pour mener son action adaptative. Ici encore, la représentation y a toute sa place. Ainsi, « l'instrument constitue pour le sujet un ensemble de contraintes qui s'imposent à lui et qu'il doit gérer dans la singularité de chacune de ses actions » (Rabardel, 1995). L'individu cherche donc à faire appel aux schèmes constitutifs de ses représentations pour pouvoir s'approprier l'instrument. En effet, le processus d'instrumentalisation concerne à la fois l'artefact et les schèmes sociaux d'utilisation. Il y a, donc une appropriation de l'artefact d'une part et l'adaptation des schèmes privés d'autres part. (Rabardel et Samurçay, 2006). Ainsi, tout comme lors de sa confrontation à un réel inconnu qu'il transforme en connu au moyen de l'adaptation des schèmes périphériques de sa représentation, l'individu, dans une volonté d'appropriation de l'outil va d'abord devoir se confronter à l'artefact. La genèse instrumentale (Rabardel, 1995) introduit l'idée que tout outil doit être avant tout pensé, afin de mener à bien une action instrumentée. Dans cette dynamique, l'individu face à l'artefact mobilise ses représentations, afin de donner sens à ce nouvel élément, afin de l'instrumenter. Cette instrumentation constitue en quelques sortes une appropriation personnelle de l'artefact. Cela illustre les multiples fonctions des représentations et leur influence sur l'action de l'individu. Ainsi l'individu va spécifiquement instrumenter, conceptualiser une partie de l'artefact, pour répondre à ses besoins par la mobilisation de ses représentations (Voir Fig. 3).

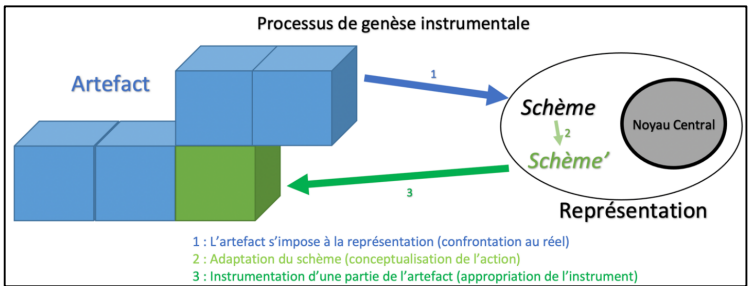


Fig. 3. Modèle de la genèse instrumentale sous le prisme de la représentation (Ajouaou, 2024)

4 Méthodologie de l'étude empirique menée auprès des cadres de santé formateurs

4.1 Contexte de l'étude et caractéristiques de la population cible

Cette étude s'est intéressée à l'identification et la cartographie des éléments constitutifs de la représentation sociale d'un panel de 20 cadres de santé formateurs de l'AP-HP répartis sur 4 IFSI différents. Chacun des IFSI sollicités au cours de cette étude a traversé la pandémie de la covid 19, et la mise en place d'une hybridation de la formation

professionnelle afin d'assurer la continuité pédagogique durant le confinement. Trois ans après la pandémie, nous avons cherché à analyser la représentation des cadres de santé formateurs rencontrés (Tableau 1). A dominance féminine (3 hommes pour 17 femmes), le panel étudié est relativement représentatif de la population soignante en générale. Sur les 20 cadres rencontrés travaillant dans quatre IFSI, 16 occupaient un poste de formateur et 4 occupaient la mission de coordination pédagogique de la structure. Enfin, parmi ces 20 cadres de santé, la moitié avaient entre 45 et 59 ans, 7 cadres entre 44 et 24 ans et 3 cadres avaient 60 ans ou plus.

Tableau 1. Caractéristiques de la population cible

Caractéristiques		Effectif N= 20
Sexe	Homme	3
	Femme	17
Age	24-44ans	7
	45-59ans	10
	60ans et plus	3
Poste occupé	Formateurs	16
	Coordination pédagogique	4

4.2 Méthodologie de recueil et analyse des données

La méthode de recueil de données relative à cette étude a été basée sur des entretiens semi-directifs à l'aide d'une grille d'entretien. Nous appuyant sur les travaux de Guimelli (1995) et Moliner et Guimelli (2015), nous avons structuré notre guide d'entretien autour de trois fonctions essentielles de la représentation. En effet, nous avons cherché à mettre en avant dans la structure même de cette grille d'entretien les fonctions descriptive, prescriptive et évaluatrice de la représentation. De plus, pour réaliser une analyse prototypique de la représentation des cadres de santé formateurs, nous nous sommes appuyés également sur les travaux de Vergès (1992) et Abric (2003a). Ainsi, nous avons demandé aux participants, en début d'entretien, de donner 4 mots relatifs à la formation hybride et au cours de l'entretien nous leur avons ensuite demandé de donner le rang d'importance de ces 4 propositions. Au cours de l'analyse prototypique, nous avons calculé les rangs moyens d'importance par le biais d'un score d'importance ainsi que la fréquence moyenne d'apparition des mots spécifiques choisis par les cadres de santé tout au long de l'entretien afin d'en repérer la centralité (Vergès, 1992 ; Abric, 2003a) et d'en dresser la classification de ces derniers.

5 Principaux résultats de l'étude

5.1 Les mots décrivant l'hybridité

Cette première collecte par évocation hiérarchisée nous a permis d'identifier 48 mots à analyser (Tableau 2). A partir de ces données, nous avons croisé deux critères possibles de « prototypicalité » (Vergès, 1992). Tout d'abord l'importance que les cadres de santé formateurs donnaient à leurs propres mots en analysant les rangs moyens d'importance. De plus, nous avons réalisé le calcul de la fréquence de ces mots dans les propos des formateurs interrogés au cours de nos entretiens en calculant la fréquence moyenne et le rang moyen d'importance à partir d'un score d'importance de chacun des 48 mots.

Tableau 2. Résultat de l'évocation hiérarchisée des 20 formateurs rencontrés

Les 48 Mots issues de l'évocation hiérarchisée							
1	Complexité	13	Asynchrone	25	Moyens	37	Variété
2	Binarité	14	Faisabilité	26	Partage	38	Concordance
3	Autonomie	15	Moodle	27	Outil-numérique	39	Pédagogie
4	Image	16	Numérisation	28	Dynamique	40	Viso-conférence
5	Fourre-tout	17	Praticité	29	Apprentissage	41	Nouveauté
6	Modernité	18	Ordinateur	30	Digitalisation	42	Présentiel
7	Flou	19	Diversifier	31	Appropriation	43	Alternance
8	Formation	20	Compétence	32	Equipement	44	Organisation
9	Temps	21	Synchrone	33	Simulation	45	Accompagenement
10	Internet	22	Technicité	34	Learning-apps	46	Mélange
11	Outils	23	Feed-back	35	Covid	47	Numérique
12	Mixte	24	Université / FAC	36	Méthodes	48	Distanciel

Respectant la méthodologie prototypique (Abric, 2003b), nous avons mis en lumière la centralité de 10 mots comme organisateurs de la représentation qu'ont les cadres de santé formateurs d'un dispositif de formation hybride (Tableau 3). Concernant la fréquence d'apparition, 12 mots se sont avérés avoir une fréquence moyenne d'apparition supérieure à la moyenne (Tableau 4).

Tableau 3. Distribution des mots ayant un rang moyen d'importance supérieure à la moyenne

MOTS	Rang moyen d'importance	Moyenne
DISTANCIEL	1,35	0,2
NUMERIQUE	0,75	
MELANGE	0,5	
ALTERNANCE, ORGANISATION, ACCOMPAGNEMENT	0,4	
PRESENTIEL	0,35	
VISIO-CONFERENCE, NOUVEAUTE	0,3	
PEDAGOGIE	0,25	

Tableau 4. Distribution des mots ayant une fréquence moyenne d'apparition supérieure à la moyenne

MOTS	Fréquence moyenne d'apparition	Moyenne
DISTANCIEL	28,85	5,1
FORMATION	25,4	
TEMPS	22,89	
PRESENTIEL	19,15	
OUTILS	18,5	
NUMERIQUE	17,6	
PEDAGOGIE	15,65	
UNIVERSITE / FAC	10,7	
SYNCHRONE	6,35	
ASYNCHRONE	5,7	
ORDINATEUR, OUTIL-NUMERIQUE	5,45	

5.2 La cartographie du noyau central de la représentation

Dans cette logique descriptive (Abric,2003b) il nous a été possible de cartographier la représentation qu'ont les cadres de santé formateurs de l'hybridité de la formation professionnelle (Tableau 5).

Tableau 5. Analyse prototypique, la structure des représentations (Abric, 2003b)

		Rang Moyen d'importance	
		Importance Haute. *> à la moyenne*	Importance Faible. *< ou = à la moyenne*
Fréquence moyenne	Fréquence Haute. *> à la moyenne*	Noyau central	1ère Périphérie
	Fréquence Base. *< ou = à la moyenne*	Eléments contrastés	2e Périphérie

La case supérieure gauche abrite les éléments les plus fréquemment cités (fréquences moyennes supérieures à la moyenne) et ayant des rangs moyens d'importance plus hauts (rang moyen d'importance supérieur à la moyenne), qui sont constitutifs du noyau central. Les éléments périphériques se trouvent, quant à eux, en premier lieu dans la case supérieure droite. Il s'agit des éléments les plus fréquemment cités mais ayant un rang moyen d'importance plus faible. Ils sont alors constitutifs de la « première périphérie » (Abric, 2003b), alors que ceux de la case inferieure droite, les éléments les moins cités et les moins importants constituent la « seconde périphérie » (Abric, 2003b). Les éléments figurant dans la case inferieure gauche sont des éléments énoncés

par peu de personnes mais ces dernières leur attribuent une grande importance. Ces éléments peuvent alors être évocateurs de l'existence d'un sous-groupe dans le groupe initial, « porteurs d'une représentation différente » (Abric, 2003b). A la lecture du Tableau 6, il est possible d'observer que pour ce groupe de cadres de santé formateurs, la représentation d'un dispositif de formation hybride est organisée autour d'un noyau central à quatre éléments : un dispositif de formation hybride constitue de la « pédagogie », en « présentiel » et en « distanciel », via le « numérique ». De plus, nous constatons une première périphérie qui repose sur un réel besoin de « formation » et de « temps » afin de posséder et de manipuler les « outils », les « outils numériques » et « l'ordinateur », afin de réaliser les activités hybrides « synchrone » ou « asynchrone » dans un contexte « universitaire ». La seconde périphérie est constituée, quant à elle, d'éléments davantage en lien avec le contexte. En effet, la pandémie de la « Covid » a impulsé une « dynamique » quant à la « variété » de « méthodes » d'« apprentissage » tels que la « simulation » et l'utilisation de « learning-apps », dans une logique de « digitalisation » des contenus mais aussi par « l'appropriation » de nouveaux « équipements » en « concordance » avec les nouvelles pratiques. Pour finir, il est également possible de repérer certains éléments contrastés quant à la « nouveauté » de réaliser des cours par « visio-conférence », pouvant questionner « l'organisation » de cette formation en « alternance ». Ce « mélange » de moyens et méthodes pourrait avoir un impact sur « l'accompagnement » des apprenants au sein de la formation ou bien relèverait davantage d'un besoin « d'accompagnement » à la « nouveauté ».

Tableau 6. Cartographie de la représentation des 20 formateurs rencontrés

La représentation des 20 cadres de santé sur un dispositif de formation hybride

		Rang Moyen d'importance					
		Importance Haute. *> à 0,2*			**Importance Faible.** *< ou = à 0,2*		
		Mots	**Fréquence**	**Rang**	**Mots**	**Fréquence**	**Rang**
Fréquence moyenne	**Fréquence Haute.** *> à 5*	Distanciel	28,85	1,35	Formation	25,4	0,05
		Présentiel	19,15	0,35	Temps	22,85	0,1
		Numérique	17,6	0,75	Outils	18,5	0,1
		Pédagogie	15,65	0,25	Université/FAC	10,7	0,15
					Synchrone	6,35	0,15
					Asynchrone	5,7	0,1
					Outils numériques	5,45	0,15
					Ordinateur	5,45	0,15
		Mots	**Fréquence**	**Rang**	**Mots**	**Fréquence**	**Rang**
	Fréquence Base. < ou = à 5	Accompagnement	4,5	0,4	Méthodes	4,6	0,2
		Viso-conférence	4,3	0,3	Apprentissage	4,1	0,2
		Nouveauté	3,8	0,3	Simulation	2,75	0,2
		Organisation	3,35	0,4	Covid	2,3	0,2
		Alternance	3	0,4	Learning-apps	2,2	0,2
		Mélange	2,2	0,5	Appropriation	1,75	0,2
					Digitalisation	1	0,2
					Dynamique	0,9	0,2
					Variété	0,5	0,2
					Equipement	0,4	0,2
					Concordance	0,35	0,2

6 Discussion

Cette étude visait à explorer la représentation sociale de l'hybridité qu'ont les formateurs de la formation professionnelle, tout particulièrement la formation infirmière. Nous avons choisi d'analyser la représentation sociale des cadres de santé formateurs car elle constitue une forme de connaissance partagée par un groupe professionnel. Ceci nous semblait nécessaire pour comprendre comment les formateurs de la formation professionnelle perçoivent et donnent sens à leur environnement de travail dans un contexte d'hybridation de la formation. Cette analyse permet précisément de repérer, au travers de la structure de la représentation sociale (son noyau central et ses périphéries), les éléments, significations et hiérarchisations qui sont collectivement partagés. Identifier ces schèmes communs, même s'ils sont répartis sur différents sites, a permis de comprendre les principes sous-jacents qui organisent leur pensée et guident potentiellement leurs pratiques face à l'hybridation. En s'appuyant sur le cadre théorique des représentations sociales, notamment l'approche structurale développée par Abric (2003) et Vergès (1992), complétée par une conceptualisation de l'hybridité, cette recherche a permis de cartographier la structure de cette représentation et de commencer à en comprendre les dynamiques.

L'analyse prototypique des évocations et la confrontation des données ont mis en évidence l'existence d'un noyau central, d'éléments périphériques et d'éléments contrastés. La théorie des représentations sociales postule que toute représentation est organisée autour d'un noyau central, considéré comme stable, résistant au changement, et essentiel pour définir l'objet représenté et organiser les pratiques (Moscovici, 1976 ; Jodelet, 1985 ; Abric, 2003 ; Flament, 2003 ; Kalampalikis, 2019). Dans cette étude, le noyau central de la représentation de la formation hybride chez les cadres de santé formateurs est constitué de quatre éléments : la pédagogie, le présentiel et le distanciel, ainsi que le numérique. Ces résultats confirment que la formation hybride est perçue, au plus profond de cette représentation, à travers des notions fondamentales qui structurent l'activité des formateurs. La pédagogie, identifiée comme un élément central, est qualifiée d'idéologique ou normative (Rouquette et Rateau, 1998). Sa présence au cœur de la représentation suggère qu'elle constitue un principe structurant fondamental pour ces professionnels, dont l'identité est intrinsèquement liée à la mission d'enseignement et au développement de compétences chez les apprenants. Les concepts de présentiel, distanciel et numérique, également centraux, sont davantage liés à la pratique et à l'action (visée praxique). Leur centralité reflète la réalité vécue de l'hybridation comme une combinaison de modalités, médiatisée par les outils numériques. L'analyse de ce noyau central confirme donc l'ancrage de la représentation dans le cœur du métier de formateur (la pédagogie) et dans les outils et modalités qui définissent concrètement la formation hybride (présentiel, distanciel, numérique). Autour de ce noyau stable, la représentation s'organise en périphéries, composées d'éléments plus flexibles et adaptatifs. Ces éléments sont en interface avec la réalité quotidienne et permettent l'interprétation et l'adaptation face aux situations nouvelles. La première périphérie comprend des éléments tels que la formation, le temps, les outils (généraux et numériques), et l'ordinateur. Ces termes sont directement liés aux pratiques quotidiennes des formateurs : leur

activité de formation auprès des étudiants, la gestion du temps imposée par les modalités synchrones et asynchrones, et l'appropriation des outils numériques nécessaires à l'enseignement hybride. La seconde périphérie intègre des éléments liés au contexte plus large, tels que, la dynamique engendrée, la variété des méthodes et d'apprentissage (simulation, learning-apps), la digitalisation et l'appropriation d'équipements. De plus, la présence de la pandémie de Covid-19 souligne son rôle de catalyseur dans l'accélération de l'hybridation et l'introduction forcée de nouvelles pratiques et méthodes. Ces éléments périphériques illustrent la manière dont les cadres de santé formateurs intègrent les influences externes et les nécessités d'adaptation dans leur représentation de la formation hybride, tout en préservant le noyau central axé sur les fondamentaux de la pédagogie et les modalités constitutives du dispositif. L'analyse a également révélé des éléments contrastés, caractérisés par une importance plus disparate au sein du groupe étudié. Il s'agit notamment de la nouveauté, de la visioconférence, de l'organisation de l'alternance, et de l'accompagnement (ou son besoin). Ces éléments contrastés sont particulièrement intéressants car ils peuvent signaler des divergences d'expériences, des perceptions moins uniformes, ou des aspects de la représentation encore en cours de cristallisation. Le contraste autour de la visioconférence peut refléter des vécus différents quant à sa capacité à maintenir le lien relationnel ou sa perception comme une modalité limitante. La nouveauté et l'accompagnement pourraient indiquer des stades différents d'appropriation de l'hybridation et des besoins variables en soutien, potentiellement liés au degré d'avancement du processus d'hybridation au sein des différents IFSI.

La complexité de l'hybridation en tant que combinaison d'éléments spatio-temporels, pédagogiques et communicationnels est bien appréhendée par la structure de la représentation mise en évidence. Le noyau central ancre la définition de la formation hybride dans ses composantes essentielles (pédagogie, modalités spatiales, outil numérique), tandis que les périphéries et éléments contrastés reflètent la complexité de sa mise en œuvre, son caractère processuel et évolutif influencé par le contexte (pandémie, universitarisation), l'adaptation des pratiques, et les variations inter-sites. L'idée de « tiers-lieu », un espace hybride où se mêlent différentes temporalités et spatialités, trouve un écho dans la centralité du présentiel et du distanciel, mais aussi dans les interrogations périphériques ou contrastées concernant le temps nécessaire, l'organisation, ou l'accompagnement à distance. La notion de présence à distance est également pertinente, les formateurs étant confrontés au défi de maintenir le lien relationnel essentiel à la socialisation professionnelle infirmière à travers des outils comme la visioconférence. En s'appuyant sur un cadre théorique multidisciplinaire, cette étude a mobilisé les concepts fondamentaux de la psychologie sociale pour comprendre la nature, la structure et les fonctions des représentations sociales. Parallèlement, la philosophie a permis de conceptualiser la notion complexe d'hybridité comme un tiers-lieu spatio-temporel singulier. Cette articulation théorique a rendu possible l'objectif principal de dresser une cartographie de la représentation sociale de ce groupe de professionnels vivant une forte hybridation de la formation. L'étude empirique a concrètement permis de mettre en lumière les éléments sur lesquels repose la représentation de l'hybridité. Cette cartographie structurelle, bien que riche en enseignements, doit être interprétée en tenant

compte de certaines limites. La principale réside dans la taille restreinte de l'échantillon d'entretien (20 cadres de santé formateurs), ce qui limite la généralisation des résultats à l'ensemble des formateurs de la formation professionnelle. Comprendre cette structure et les éléments partagés par le groupe professionnel a permis de mieux appréhender la réalité vécue par les professionnels évoluant dans un dispositif de formation instrumenté et en forte hybridation. La représentation sociale agit comme une grille d'interprétation du réel et guide les pratiques, influençant ainsi l'appropriation des outils numériques et l'adaptation aux nouvelles modalités d'enseignement. Cette recherche, en offrant une vision structurée de la représentation sociale de l'hybridation, ouvre d'autres perspectives d'étude. En effet, il serait particulièrement pertinent d'explorer le lien entre cette représentation et l'agentivité effective des formateurs face à cette hybridation, c'est-à-dire leur capacité à exercer un contrôle sur leur pratique et leur environnement. De plus, il serait nécessaire de questionner l'environnement de travail lui-même, en analysant comment les contraintes et ressources institutionnelles et locales interagissent avec cette représentation et la capacité des formateurs à agir et à innover dans ce contexte de mutation de la formation professionnelle.

Références

1. Abric, J. C. (2003a). 8. L'étude expérimentale des représentations sociales. Dans : Denise Jodelet éd., Les représentations sociales (pp. 203-223). Paris cedex 14 : Presses Universitaires de France. https://doi-org.ezproxy.u-paris.fr/10.3917/puf.jodel.2003.01.0203
2. Abric, J. C. (2003b). La recherche du noyau central et de la zone muette des représentations sociales. Méthodes d'étude des représentations sociales, 296.
3. Ajouaou, J. (2024). Mémoire Master 2. *La representation d'une formation hybride: Le cas des cadres de santé formateurs.* Faculté PsySEF-Université de Lille.
4. Baudrit, A. (2018). Le tutorat universitaire à distance : examen d'une méthode basée sur la communication médiatisée par les TIC. Revue française de pédagogie. Recherches en éducation, (202), 117-138.
5. Bergson Henri, 1922. Durée et simultanéité, Paris, Alcan.
6. Dubar, C. (2008). Temporalité, temporalités: philosophie et sciences sociales. Temporalités. Revue de sciences sociales et humaines, (8).
7. Durkheim, E. (1898). Représentations individuelles et représentations collectives. Revue de métaphysique et de morale. Vol. VI, pp. 273-302.
8. Doise, W. (1990). Les représentations sociales. Traité de psychologie cognitive, 3(111-176).
9. Flament, C. (2003). 9. Structure et dynamique des représentations sociales. Dans : Denise Jodelet éd., Les représentations sociales (pp. 224-239). Paris cedex 14: Presses Universitaires de France. https://doi-org.ezproxy.u-paris.fr/10.3917/puf.jodel.2003.01.0224
10. Flichy, P. (2021). Postface. « Une sociologie de l'hybridité ». Dans : Olivier Martin éd., Les liens sociaux numériques (pp. 287-299). Paris : Armand Colin. https://doi-org.ezproxy.u- paris.fr/10.3917/arco.marti.2021.01.0287
11. Foucault, M. (1967). « Des espaces autres ». Empan, no<(sup>54), 12-19.2004 https://doi-org.ezproxy.u-paris.fr/10.3917/empa.054.0012
12. Guimelli, C. (1995). Valence et structure des représentations sociales.Bulletin de psychologie, 49(422), 58-72.
13. Halpern, G. (2020). Tous centaures ! éloge de l'hybridation. Paris : Ed. Le Pommier
14. Jacquinot, G. (1993). Apprivoiser la distance et supprimer l'absence ? ou les défis de la formation à distance. Revue française de pédagogie, 55-67.

15. Jézégou, A. (2019). Traité de la e-formation des adultes. De Boeck Supérieur.
16. Jézégou, A. (2022a). La présence à distance en e-formation : enjeux et repères pour la recherche et l'ingénierie. Presses Universitaires du Septentrion.
17. Jurgenson, N. (2012). When atoms meet bits: Social media, the mobile web and augmented revolution. Future internet, 4(1), 83-91.
18. Jodelet, D. (1985). Représentation sociale: Phénomènes, concept et théorie. In : MOSCOVICI, Serge, Psychologie sociale. Paris : PUF. Quadrige Manuels.
19. Kalampalikis N (2019). Serge Moscovici : Psychologie des représentations sociales. ⟨hal- 02091985⟩
20. Leibniz, G. W. (1962). 1704. Nouveaux Essais sur l'entendement humain.
21. Moliner, P., Guimelli, C. (2015). Les représentations sociales. Presses universitaires de Grenoble. https://doi-org.ezproxy.u-paris.fr/10.3917/pug.guime.2015.01
22. Morin, E. (1990). Introduction à la pensée complexe. Paris : Ed. Du Seuil.
23. Moscovici, S. (1976) : Psychologie des représentations sociales. In Revue européenne des sciences sociales et Cahiers Vilfredo Pareto, 14(38-39), 409-416. https://hal.univ-lyon2.fr/hal- 02091985
24. Mugny, G., & Carugati, F. (1989). Social representations of intelligence. Cambridge University Press.
25. Paquelin, D. (2011). La distance : questions de proximités. Distances et savoirs, 9(4), 565-590.
26. Piaget, J. & Inhelder, B. (1966), La psychologie de l'enfant, PUF, Paris.
27. Proulx S. (1994). « Communication publique, identité culturelle et rapports sociaux », Recherches sociographiques, vol. 35, no 1, 1994, p. 87-96.
28. Quéau, P. (1993). La pensée virtuelle. Réseaux, 61, 67-78. https://www.cairn.info/revue--1993- 5-page-67.htm.
29. Rabardel, P. (1995). Qu'est-ce qu'un instrument. Les dossiers de l'Ingénierie éducative, 19, 61- 65.
30. Rabardel, P. & Samurçay, R. (2006). De l'apprentissage par les artefacts à l'apprentissage médiatisé par les instruments. Dans : Jean-Marie Barbier éd., Sujets, activités, environnements : Approches transverses (pp. 31-60). Paris cedex 14 : Presses Universitaires de France. https://doi.org/10.3917/puf.barbi.2006.01.0031
31. Ricoeur, P. (1984). L'idéologie et l'utopie : deux expressions de l'imaginaire social. Autres temps, 2(1), 53-64.
32. Rouquette, M. L., & Rateau, P. (1998). Introduction à l'étude des représentations sociales. Presses universitaires de Grenoble.
33. Varga, R. (2013). Rapport au temps et orchestration des temporalités en formation. Distances et médiations des savoirs. Distance and Mediation of Knowledge, 1(2).
34. Vergès, P. (1992). L'évocation de l'argent : une méthode pour la définition du noyau central d'une représentation. Bulletin de psychologie, 45(405), 203-209.
35. Vial, S. (2016). La fin des frontières entre réel et virtuel : vers le monisme numérique.
36. Witmer, BG et Singer, MJ (1998). Mesurer la présence dans des environnements virtuels : un questionnaire de présence. Présence , 7 (3), 225-240.

Analyse de l'impact d'une application de jeu sérieux sur les élèves des lycées d'enseignement agricole

Luis Alberto Pinos Ullauri[1[0000−0002−3694−3487]], Catherine Pons[1,2[0000−0003−0237−3022]], Azzeddine Benabbou[1[0000−0001−8445−1770]] et Julien Broisin[1[0000−0001−8713−6282]]

[1] IRIT, Université de Toulouse, Toulouse, France
[2] INU Champolion, Albi, France

Résumé Ces dernières années, la popularité des jeux sérieux a encouragé la recherche et leurs transpositions dans plusieurs contextes éducatifs. Cependant, à notre connaissance, il existe peu de jeux sérieux destinés à l'enseignement agricole. De plus, les effectifs de l'enseignement agricole ont progressivement diminué au cours des dix dernières années. Pour endiguer ce phénomène, les jeux sérieux représentent une double opportunité car ils permettent d'augmenter l'attractivité auprès des élèves et d'accélérer la transformation numérique des établissements. Notre étude est centrée sur l'impact d'un jeu sérieux destiné aux lycées professionnels agricoles et interroge les variables suivantes : sentiment d'auto-efficacité, motivation, utilisabilité et expérience utilisateur. Une expérimentation a été menée auprès de quatre lycées professionnels agricoles répartis sur le territoire français. Les résultats montrent qu'il existe une variation significative entre le sentiment d'auto-efficacité perçu par les élèves avant et après l'utilisation du jeu sérieux. De plus, nous explorons le cas des élèves dont le sentiment d'auto-efficacité a subi une forte variation (avant versus après l'expérience). Nous avons aussi mesuré l'utilisabilité, l'expérience utilisateur et analysé les corrélations entre ces variables.

Mots-clé: Jeux sérieux · Enseignement professionnel agricole · Motivation · Auto-efficacité · Expérience utilisateur · Utilisabilité

Abstract In the last decade, the popularity of Serious Games has promoted both its research and application in several educational contexts. However, their application in the context of agricultural education has not been sufficiently explored, to the best of our knowledge. Moreover, agricultural education has seen a decrease on the number of students in the last 10 years. With the aim to increase the attractiveness of agricultural education, we propose the use of Serious Games. Consequently, we study the impact of a serious game adressed to technical agricultural high schools and to assess its impact on the following variables: self-efficacy, motivation, usability, and User Experience. This paper intents on exploring these impacts based on data collected from four technical agricultural secondary schools in France. The results show that there

is a significant variation between the perceived self-efficacy of the students before and after they participated in the experiment. Moreover, we explore the cases of students that perceived their self-efficacy considerable different before and after the experiment. We also measured the usability, the user experience and analysed the correlations between these variables.

Keywords Serious Games· Technical Agricultural Education · Motivation · Self-Efficacy · User Experience · Usability

1 Introduction

De nos jours, il est de plus en plus courant de former les futurs professionnels à l'aide d'environnements informatiques pour l'apprentissage humain (EIAHs) tels que les jeux sérieux. Bien qu'il y ait débat sur la définition des jeux sérieux [4], la communauté s'accorde sur une vision de jeux dont l'objectif n'est pas un simple divertissement [2, 3, 13, 4]. En général, ces jeux visent un objectif marketing, politique, éducatif, de sensibilisation ou autre [10]. Leur contenu est utilisable en dehors du cadre vidéo-ludique. Dans le contexte éducatif, leur contenu comprend des contenus pédagogiques, des mécanismes vidéo-ludiques inspirés de jeux vidéos et des outils de surveillance facilitant l'affichage automatisé d'un debriefing en fin de partie [27, 32]. Ils peuvent également inclure divers éléments tels que du feedback pédagogique, un système de support ou des systèmes d'interactions complexes.

Des jeux sérieux ont déjà été utilisés dans divers contextes, notamment dans l'enseignement professionnel technique, y compris dans le domaine agricole [16, 30, 17]. Les travaux de [17, 16] se sont basés sur le développement et le déploiement d'un jeux sérieux pour aider les fermiers dans la production bovine en Colombie. Un système d'aide à la décision a été mis en œuvre pour faciliter la gestion de la production. Pour cela, les auteurs se sont appuyés sur un système simulé utilisant des variables et contraintes réalistes. L'étude menée par [30] a proposé deux jeux sérieux, qui combinent les cadres des EIAHs avec les mécaniques ludo-pédagogiques, pour enseigner comment bien entretenir des plantes.

Alors que le réseau d'enseignement agricole est un élément vital pour la durabilité et la croissance économique de notre société [24], le nombre d'élèves en formation initiale dans l'enseignement agricole a baissé de 11% en dix ans [11]. Le rapport sur l'urgence d'une transition agro-politique a souligné le besoin d'agir et de renforcer l'attractivité de l'enseignement agricole auprès de potentiels élèves et de leur famille. De plus, le rapport d'orientation des élèves vers l'enseignement agricole [8] décrit des stratégies pour améliorer la communication externe de l'enseignement agricole, et ainsi attirer plus d'élèves dans ces formations. Une des propositions consiste à adapter l'enseignement professionnel à la réalité du terrain et à le mettre à la pointe du numérique. Ainsi, le projet Agr-eSchool ambitionne de concevoir des jeux sérieux d'apprentissage composés d'activités numériques basées sur des scénarios interactifs et/ou immersifs créés à partir de situations professionnelles réelles. Ces jeux sérieux s'adressent exclusivement à

l'enseignement secondaire agricole [15] et sont développés par deux sociétés du secteur EdTech pour le réseau d'établissements professionnels agricoles français UNREP.

Espinosa et García [14] ont procédé à une analyse systématique de la littérature, à partir de laquelle ils ont comparé des études centrées sur l'application de jeux sérieux dans l'enseignement agricole. Ils ont souligné l'importance de créer des jeux avec des défis agricoles spécifiques. Suivant ces recommandations, le jeux sérieux développé dans le cadre du projet Agr-eSchool propose aux élèves des situations scénarisées proches du contexte réel et ambitionne d'attirer les élèves vers l'enseignement agricole. Notre étude porte sur l'usage et l'impact de ce jeux sérieux. Pour mieux comprendre et approfondir la connaissance de l'impact des jeux sérieux sur l'enseignement agricole professionnel, cet article vise à analyser leurs effets sur le sentiment d'auto-efficacité des élèves et leur motivation. Pour cela, nous complétons notre approche par une évaluation de l'utilisabilité et l'expérience utilisateur des élèves inscrits dans les établissements participants. Dans cet objectif, nous tentons de répondre aux questions de recherche suivantes :

- **QR1 :** L'introduction d'un jeu sérieux dans un cursus professionnel a-t-elle un impact sur le sentiment d'auto-efficacité des apprenants ? Si oui, quels sont les profils d'apprenants pour lesquels l'impact est le plus important lorsque l'on considère leur motivation et la fréquence avec laquelle ils jouent à des jeux vidéos ?
- **QR2 :** Quelle est l'utilisabilité et la qualité de l'expérience utilisateur perçues du jeux sérieux ?
- **QR3 :** L'utilisabilité et l'expérience utilisateur perçues du jeux sérieux sont-elles corrélées avec l'auto-efficacité et la motivation des apprenants ?

La section suivante présente la méthodologie mise en œuvre. Elle décrit les caractéristiques principales du jeu sérieux, le protocole expérimental mis en place pour répondre aux questions de recherche proposées, les données collectées et les instruments de mesure. La section 3 décrit les résultats obtenus à partir de chacune des variables d'intérêt. Ensuite, nous discutons les résultats, limitations et perspectives de l'étude avant de conclure.

2 Méthodologie

2.1 Le Jeu Sérieux : Agr-eSchool

Dans le jeu de l'oie traditionnel, les joueurs sont engagés dans une course qui repose sur des lancés de dés. Contrairement à la version traditionnelle, le premier plateau de jeu du projet Agr-eSchool représente une version numérique et collaborative du jeu de l'oie. Il offre aux joueurs la possibilité de progresser en équipe non pas en lançant un dé, mais plutôt en collectant des points obtenus à travers la réalisation d'activités pédagogiques. La partie se joue en équipe, et l'équipe est constituée de l'ensemble des élèves d'une même classe. La mission

consiste à atteindre ensemble le sommet d'un pic montagneux. Les cases du plateau de jeu symbolisent le chemin à emprunter pour atteindre le sommet tandis que la dernière case représente le but ultime et la fin du jeu (voir la Figure 1).



Fig. 1. Capture d'écran du plateau de jeu, qui demande aux élèves de passer par toutes les cases jusqu'au sommet de la montagne.

Des activités pédagogiques ou défis éducatifs sont placés sur des cases spécifiques. Une fois réalisés, les joueurs obtiennent des points de déplacement qui leur permettent de progresser sur le chemin. Le nombre de points attribués dépend de la rapidité avec laquelle le joueur fournit la bonne réponse. De plus, les points gagnés peuvent débloquer des jokers utilisables tout au long de la partie pour aider les apprenants dans certains défis. La Figure 2 présente les modalités de progression dans le jeu.

Le contenu pédagogique comprend non seulement des capsules vidéos et des animations, mais aussi des éléments tels que des questions vrai-faux, des questions à choix multiples et des exercices d'association par glisser-déposer. La Figure 3 présente un exemple d'exercice pédagogique d'association, où les apprenants doivent associer des éléments après avoir regardé des vidéos et écouté des enregistrements. Ce contenu s'appuie sur des contextes très proches de la réalité. Dans cet exemple, les élèves doivent décrire l'ordre des étapes à suivre pour réaliser un marcottage aérien.

En outre, tous les modules pédagogiques du parcours sont destinés à préparer les élèves pour l'épreuve finale. Cette épreuve s'inspire du jeu du pendu, dans lequel on doit deviner un mot caché en proposant des lettres composant ce mot. Le mot à trouver est en rapport avec le contenu pédagogique consulté durant la partie (voir la Figure 4). Dans cette épreuve, tous les apprenants peuvent utiliser

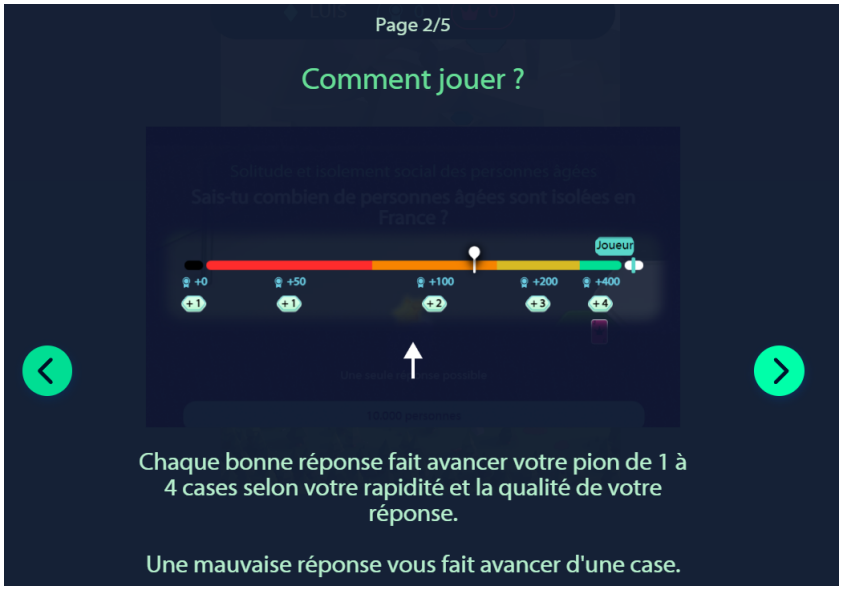


Fig. 2. Modalités de progression dans le jeu

les jokers obtenus au cours de la partie et ainsi obtenir des indices facilitant la résolution du jeu du pendu.

2.2 Protocole expérimental

L'UNREP, le partenaire fédérateur des établissements d'enseignement agricole, a sollicité les lycées professionnels de son réseau pour expérimenter le jeu sérieux et les scénarios interactifs développés dans le cadre du projet. Les expérimentations ont impliqué des élèves répartis dans 4 lycées professionnels agricoles localisés sur le territoire français. Les élèves ont rempli, avant les séances expérimentales, un premier questionnaire composé de questions socio-démographiques et du questionnaire standardisé *Academic Motivation Scale* (AMS) [31] utile pour mesurer les différents types de motivation dans les études secondaires. Ensuite, le jour de la séance, les élèves ont rempli un deuxième questionnaire standardisé *Motivated Strategies for Learning Questionnaire* [26] (MSLQ) comprenant des items pour mesurer leur sentiment d'auto-efficacité. Puis ils ont utilisé le jeu dans un contexte écologique en présence de leur enseignant. Deux séances d'apprentissage au travers du jeu sérieux se sont déroulées le même jour. Lors de la première séance, une phase de prise en main consistait à se familiariser avec les mécaniques de jeu. Ensuite, les élèves ont utilisé au travers du jeu des contenus pédagogiques spécifiquement ciblés par l'enseignant. Dans chacune des séances, les élèves ont joué sur deux scénarios distincts. Pour finir, ils ont renseigné un questionnaire post-test composé des questionnaires standardisés MSQL, *System*



Fig. 3. Capture d'écran du contenu pédagogique

Usability Scale (SUS) et Attrakdiff utiles pour évaluer respectivement leur sentiment d'auto-efficacité, l'utilisabilité du logiciel, et leur expérience utilisateur.

Les enseignants ayant participé aux expérimentations se sont tous portés volontaires dans les lycées candidats. Il y a donc une certaine motivation des lycées à expérimenter les jeux dans les cursus pédagogiques, mais nous n'avons pas dispensé de questionnaires spécifiques à destination des enseignants pour recueillir leur motivation.

2.3 Données collectées et instruments

La population d'apprenants ayant participé aux expérimentations correspond aux élèves de l'ensemble des établissements ayant répondu au premier appel lancé par l'UNREP. La population des élèves en lycée agricole est classiquement relativement hétérogène [9]. Ici, elle est composée de 35 élèves (Age : Min = 13, Max = 20, Moy = 16.09, ET = 1.36), dont 16 femmes (45.71 %) et 19 hommes (54.29 %). Le niveau des élèves est également diversifié, avec 42.86% en première année de BAC Professionnel, 25.71% en deuxième année, et le reste en CAP Agricole (28.57% en première et 2.86% en deuxième). Au niveau des filières, la majorité (40%) des élèves proviennent de la filière "Service à la personne", tandis que près de 54% sont issus de métiers liés à l'agriculture tels que les métiers de l'agroéquipement, de l'horticulture et des aménagements paysagers par exemple. En termes de conditions et troubles, la majorité affirme ne pas avoir de handicaps. Cependant, il y a quelques cas de troubles de déficit de l'attention et hyperactivité (TDAH), de dyslexie et de dysphasie. Ces premières descriptions renforcent l'idée que le public de ce type d'établissements est hétérogène et diversifié. Ceci ne remet pas en cause la validité des analyses mais permet d'avoir une vision plus large qui reflète davantage la réalité du terrain.



Fig. 4. Dernier item du jeu : La devinette

Motivation des élèves. D'après [31], un des facteurs les plus importants dans l'éducation est la motivation. Ils ont développé et validé l'Échelle de Motivation en Éducation qui a pour but de mesurer la motivation sous différentes dimensions : la motivation intrinsèque à la connaissance ; la motivation intrinsèque à l'accomplissement ; la motivation intrinsèque à la stimulation ; la motivation extrinsèque identifiée ; la motivation extrinsèque introjectée ; la motivation extrinsèque régulation externe ; et finalement, la démotivation. Leur questionnaire comprend un total de 28 items, avec 4 items dédiés à chaque dimension. Ces items sont évalués à l'aide d'une échelle de Likert à 5 niveaux allant de *Pas du tout d'accord* à *Complètement d'accord*. Cet outil de mesure pré-expérimental a été proposé aux participants avant l'utilisation du jeu sérieux.

Auto-efficacité des élèves. Le sentiment d'efficacité personnelle, un concept introduit par Bandura en 1977 [5, 6], joue un rôle clé dans l'engagement de l'apprenant. Des personnes avec des aptitudes identiques peuvent avoir des performances différentes selon le niveau de leur sentiment d'efficacité [23]. Par conséquent, le sentiment d'auto-efficacité peut influencer non seulement la performance, mais aussi la persévérance et la motivation des élèves, qui sont essentielles à l'apprentissage. Les travaux de [25] se sont focalisés sur la recherche des techniques utilisées pour soutenir ce sentiment vis-à-vis des élèves. Les jeux

d'apprentissage, de par leur interactivité mais aussi les contenus pédagogiques qu'ils embarquent, peuvent affecter ce sentiment. Dans notre étude, le questionnaire, développé et testé par [26], a été utilisé pour mesurer l'auto-efficacité des élèves des lycées agricoles. Ce questionnaire est composé de 8 items, avec une échelle de Likert à 7 niveaux allant de *Pas du tout d'accord* à *Complètement d'accord*. Il a été dispensé aux élèves avant et après l'utilisation du jeu sérieux.

Utilisabilité du jeu sérieux. L'utilisabilité peut être définie comme étant le degré selon lequel un produit (le jeu sérieux dans notre cas) peut être utilisé avec efficacité, efficience et satisfaction par des utilisateurs [21]. L'outil le plus utilisé pour mesurer l'utilisabilité d'un système, le SUS (*System Usability Scale*), a été proposé en 1996 [12]. Les travaux de [1] ont permis de vérifier la validité et la robustesse de l'outil.

Le questionnaire est composé de 10 éléments mesurés par une échelle de Likert à 5 niveaux allant de *Pas du tout d'accord* à *Entièrement d'accord*. Les résultats du questionnaire permettent de calculer un score sur 100 qui peut être interprété selon des valeurs de référence [12, 29]. Le questionnaire SUS, traduit en français, a été délivré aux élèves après l'utilisation du jeu.

Expérience Utilisateur. Le travail de [20] souligne que l'expérience utilisateur est plus large que la notion d'utilisabilité et peut englober cette dernière, qui a certainement un effet sur l'expérience de l'utilisateur. Le travail de [22] s'est basé sur la traduction et la validation de la version française de l'outil AttrakDiff 2 [19]. D'après ce modèle, les utilisateurs perçoivent les produits interactifs selon deux dimensions : (1) les qualités pragmatiques, qui se concentrent sur le produit (c'est-à-dire qu'un produit perçu comme structuré ou clair est considéré comme un produit avec de bonnes qualités pragmatiques) ; (2) les qualités hédoniques, qui se concentrent plutôt sur le soi (par exemple, un produit perçu comme original ou captivant aurait de bonnes qualités hédoniques). L'outil est constitué de 28 paires de *mots contrastés*, dans une échelle de Likert à 7 niveaux. La version abrégée du questionnaire composée de 10 items est celle que nous avons utilisée [18]. Elle a été proposée aux élèves après l'utilisation du jeu sérieux.

2.4 Analyse de données

Les données ont été traitées et analysées dans le langage de programmation R et les libraires dplyr, stringi, plotly et ggplot2.

QR1. Pour répondre à **QR1** concernant l'impact du jeu sérieux sur le sentiment d'auto-efficacité des apprenants, nous avons mesuré les variations entre les réponses au questionnaire d'auto-efficacité avant l'utilisation du jeu sérieux, et après son utilisation. Nous avons utilisé le test-t par paires pour inférer s'il y avait bien une variation significative de ce sentiment. Ensuite, nous avons tenté d'utiliser des modèles de régression linéaire pour quantifier les liens entre la variation de l'auto-efficacité et le profil des étudiants. Cependant, le faible nombre

d'élèves ne nous a pas permis de faire converger les modèles. Par conséquent, nous avons étudié le cas particulier des élèves dont le sentiment d'auto-efficacité a subi une forte variation (avant versus après l'expérience).

QR2. Pour répondre à **QR2** concernant l'utilisabilité et l'expérience utilisateur, nous avons analysé les réponses aux questionnaires SUS et Attrakdiff. Pour le SUS, nous avons calculé le score d'après la formule proposée par [29].

QR3. Enfin, pour répondre à **QR3** concernant des liens éventuels entre les différentes variables précédemment étudiées, nous avons calculé les corrélations linéaires de Pearson entre les dimensions de l'expérience utilisateur et le score SUS de l'utilisabilité. Ces corrélations ont été réalisées à partir de chaque paire de variables. Des étoiles sont utilisées pour présenter visuellement les valeurs-p.

3 Résultats

3.1 Motivation

La Figure 5 présente un diagramme en boîtes des différentes dimensions de la motivation. Nous pouvons remarquer que la longueur des boîtes est comparable parmi les dimensions de la motivation. Nous constatons également que la démotivation est, en moyenne, la plus faible, suivie par la motivation intrinsèque à la stimulation. De plus, nous pouvons constater aussi, qu'en moyenne, la motivation extrinsèque est plus élevé que la motivation intrinsèque.

Maintenant que nous avons décrit brièvement la motivation éducative des participants à l'expérimentation, nous présentons dans les sous-sections suivantes les résultats des analyses réalisées pour répondre à nos questions de recherche.

3.2 Résultats QR1

De manière générale, les réponses aux items du questionnaire du sentiment d'auto-efficacité ont des tendances similaires. Dans le but de visualiser ces tendances, la Figure 6 présente l'histogramme de l'item 4 : « J'ai confiance que je peux comprendre le contenu le plus difficile présenté par les enseignants dans cette matière». Nous remarquons une légère diminution des niveaux les plus bas de l'échelle, et une augmentation relativement faible des derniers niveaux, ce qui peut suggérer que les élèves ont un sentiment d'auto-efficacité plus fort après avoir utilisé le jeu sérieux.

Nous avons calculé la variation du sentiment d'auto-efficacité entre les questionnaires pré-test et post-test et normalisé les valeurs. Nous avons vérifié sa distribution de probabilité comme gaussienne avec le test de normalité de Shapiro (valeur-p = 0.93, c'est-à-dire qu'une valeur-p inférieure à 0,05 signifierait que nous devrions rejeter l'hypothèse selon laquelle la distribution n'est pas normalement distribuée). Cela nous a permis d'effectuer un test-t par paires, qui

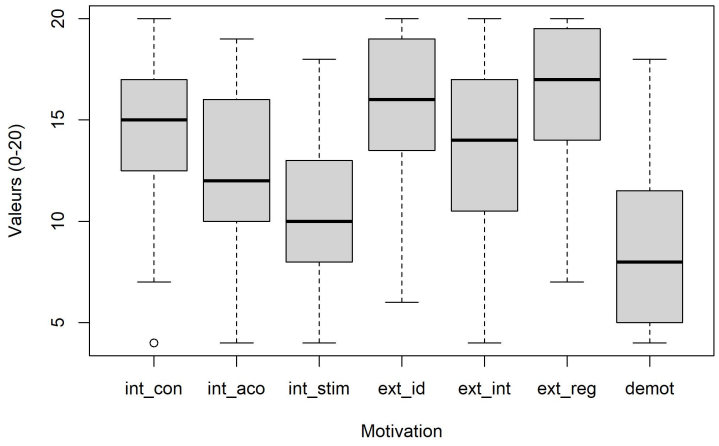


Fig. 5. Évaluation de la motivation dans les études.
Note. Les variables qui commencent par *int* représentent la motivation intrinsèque. Les variables qui commencent par *ext* représentent la motivation extrinsèque. Finalement, *demot* représente la démotivation académique.

a confirmé l'existence d'une différence significative entre le pré-test et le post-test (valeur-p = 0.02) : la différence moyenne du sentiment d'auto-efficacité de l'échantillon est égale à +3.23.

En n'ayant que 35 points de données pour l'instant, nous n'avons pas pu faire converger des modèles de régression dans l'objectif d'expliquer la variation du sentiment d'auto-efficacité en fonction de la motivation et l'expérience de jeux vidéos. La Table 1 présente les différents modèles de régression, les coefficients des prédicteurs, et les coefficients de détermination. Nous pouvons regarder les coefficients de détermination qui ne dépassent pas 6%.

De plus, nous avons analysé les cas particuliers des élèves qui ont perçu un changement du sentiment le plus bas et le plus haut d'efficacité. Nous avons pu constater quelques caractéristiques. Par exemple, trois élèves avaient une haute démotivation et une haute variation positive du sentiment d'auto-efficacité. Cependant, un élève était très démotivé et avait ressenti une variation négative du sentiment d'auto-efficacité. Nous n'avons pas pu trouver d'autres caractéristiques apparentes entre la motivation intrinsèque, extrinsèque, la fréquence de jeu et l'auto-efficacité.

3.3 Résultats QR2

Le score SUS permet de mesurer l'utilisabilité perçue du jeu sérieux. Nous rappelons que ce score peut varier entre 0 et 100. Un système est classifié comme

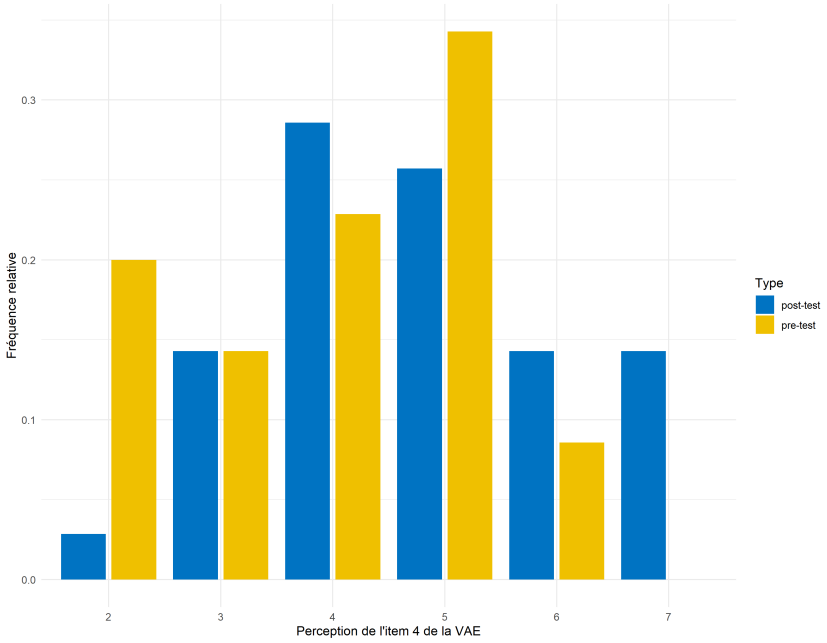


Fig. 6. Histogramme de l'item 4 du questionnaire MSLQ

Acceptable si le score est égal ou supérieur à 50.9, et Bon si le score obtenu est supérieur à 71,4 [7]. Le résultat obtenu à partir du jeu de données étudié présente une moyenne de 66.93+-5.87 (intervalle de confiance à 95%), un écart type de 17.09 et un coefficient de variation de 26%. De plus, le score minimum obtenu est de 30 et le maximum 95. La moyenne obtenue indique que le système est acceptable mais ne peut pas être qualifié comme bon.

L'AttrakDiff propose une évaluation quantitative de l'expérience utilisateur vécue au travers d'un système interactif. Il se compose d'un jeu de questions présentant des paires opposées. Les caractéristiques présentées à gauche de la Figure 7 sont considérées comme plutôt négatives alors que celles de droite sont plutôt positives. La Figure 7 présente les résultats obtenus auprès des élèves ayant utilisé le plateau 'Jeu de l'oie collaboratif' du jeu sérieux Agr-eSchool. On peut voir que les élèves ont eu un sentiment positif à divers degrés dans toutes les dimensions (hédonique et pragmatique). Du côté de la qualité pragmatique, les élèves ont perçu le jeu comme pratique, légèrement prévisible, clair et simple. En ce qui concerne l'attractivité, les élèves ont, en moyenne, trouvé le jeu beau et bon.

Tableau 1. Modèles de Régression Linéaire pour expliquer la variation de l'auto-efficacité (VAE).
Note : VD correspond à Variable dépendante. CP représente le Coefficient du Prédicteur. CD décrit le Coefficient de Détermination. Extr est l'acronyme d'Extrinsèque tandis que Intr. est l'acronyme d'Intrinsèque. EJV représente l'Expérience Jeux Vidéos

VD	Prédicteur	CP	CD
VAE	Extr. Externe	-0.02	0.5%
VAE	Extr. Introjecte	-0.05	2.9%
VAE	Extr. Identifie	-0.03	1.1%
VAE	Intr. Stimulation	-0.04	1.3%
VAE	Intr. Accomplissement	-0.03	1.4%
VAE	Intr. Connaissance	-0.07	5.2%
VAE	EJV	[0.58-EJV (2), 0.30-EJV (3), 0.39-EJV (4)]	3.6%

3.4 Résultats QR3

La Figure 8 présente la matrice de corrélation entre les variables d'intérêt, notamment la variation du sentiment d'auto-efficacité, la mesure de l'utilisabilité et les mesures de l'expérience utilisateur. De plus, les valeurs-p significatives sont marquées avec des étoiles. D'abord, toutes les corrélations linéaires sont positives à divers degrés. Nous pouvons aussi voir qu'il n'y a pas de corrélations significatives entre la différence d'auto-efficacité et les dimensions de l'expérience utilisateur ou l'utilisabilité. Par contre, il y a des corrélations significatives entre l'utilisabilité et les dimensions de l'expérience utilisateur.

4 Discussion

4.1 Discussion QR1

Les résultats montrent que la variation positive du sentiment d'auto-efficacité des élèves, après avoir joué, est statistiquement significative. De plus, les modèles de régression linéaire ne convergent pas pour l'instant. Cependant, les cas particuliers des élèves les plus démotivés nous donnent des pistes intéressantes pour analyser si le jeu sérieux peut avoir un effet plus important sur les apprenants qui, au départ, étaient plutôt démotivés. Par contre, pour bien explorer cette situation, il serait nécessaire de collecter plus de données pour mener l'analyse avec des modèles de régression mixte. Ce type d'analyse permettrait de suivre des pistes d'inférences et de causalité.

4.2 Discussion QR2

Sauro [28] a analysé 3187 réponses au questionnaire SUS pour décrire les percentiles des scores bruts obtenus. Cette étude nous permet de comparer l'utilisabilité perçue de notre jeu sérieux avec celle des autres produits de [28]. Avec un score moyen qui atteint presque 67, notre jeu sérieux se trouve dans

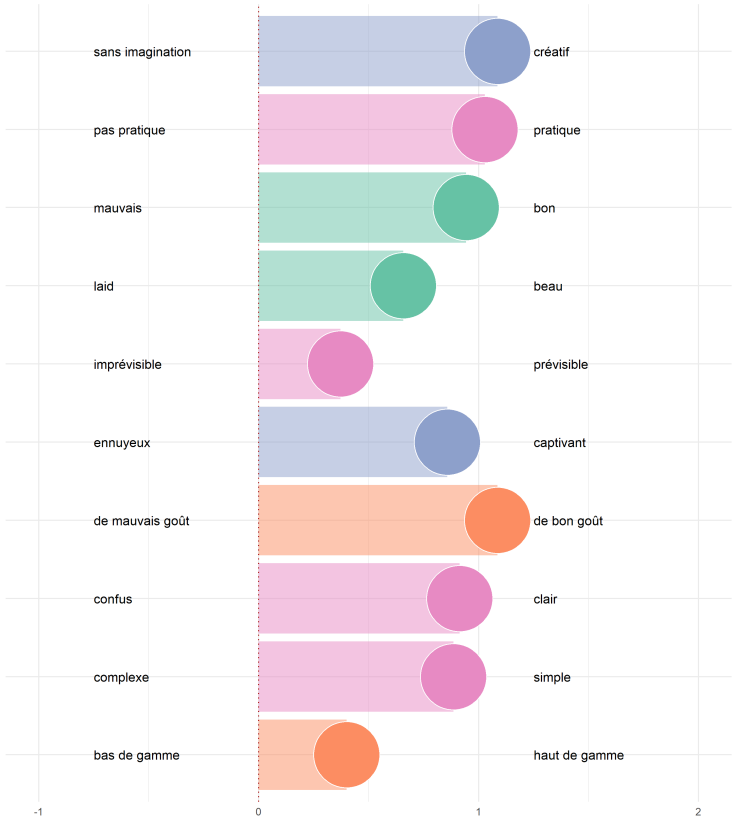


Fig. 7. Plot de l'expérience utilisateur par item et thème. Les paires de *mots contrastés* sont mis à gauche et à droite de chaque axe. Les items en vert décrivent l'attractivité, les items oranges et bleus les qualités hédoniques, et le rose la qualité pragmatique

la tranche qui correspond aux percentiles 41-59. Cela veut dire qu'entre 41% et 59% de personnes de l'analyse de Sauro auraient trouvé le jeux comme utilisable. D'après l'étude de Bangor [7], la moyenne du score SUS indique que le jeu sérieux est acceptable, et considérant la proximité de notre score vers l'échelon suivant. Cependant, il faut aussi préciser que ce type d'analyses présente des faiblesses par rapport à sa comparabilité au niveau du contexte et du public cible. Sauro [28] a aussi proposé des analyses par type d'interface, où la plus proche du jeux sérieux est celle d'un site web (puisque le jeu sérieux est basé sur un navigateur web). Avec cette analyse par type d'interface, notre moyenne (66.97) est presque identique à celle des études prises en compte (67.0) pour l'analyse des sites web. Par contre, notre écart type (17.09) est supérieur à celle des produits étudiés (13.4).

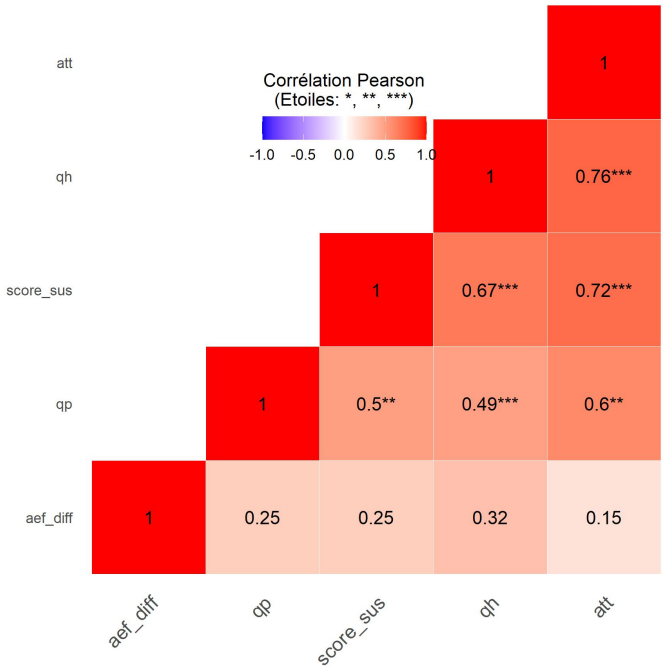


Fig. 8. Heatmap de corrélations de l'utilisabilité, l'expérience utilisateur et la variation du sentiment d'auto-efficacité
Note: ***p<0.001; **p<0.01; *p<0.05

De plus, nous constatons des moyennes positives pour toutes les qualités de l'expérience utilisateur. En général, les élèves ont perçu notre jeu sérieux comme très pragmatique. Cependant, les élèves ont une impression presque neutre dans le sens qu'ils pensent que le jeu n'est pas si prévisible. Cela a du sens car le jeu est linéaire et les objectifs sont bien mis en place dès le début de la partie. En moyenne, le jeu stimule les élèves, ce qui valide l'efficacité (au niveau stimulant) des interactions et utilisations de modules pédagogiques dans le jeu. Aussi, les élèves se sentent identifiés avec le jeu, et le trouvent de bon goût et de haute gamme. L'attractivité du jeu est aussi positive, et il est perçu, en moyenne, bon et beau.

4.3 Discussion QR3

Notre analyse de corrélations linéaires de Pearson entre les mesures de l'expérience utilisateur et l'utilisabilité montrent des valeurs de corrélation significatives pour toutes les variables, sauf pour la variation du sentiment d'auto-efficacité. Cela veut dire que pour l'instant, nous ne pouvons pas constater de corrélations entre l'impact du jeu sur l'auto-efficacité des élèves et leur perception du jeu en termes

d'utilisabilité. Néanmoins, nous pouvons constater que l'attractivité est hautement corrélée avec le score SUS et les qualités pragmatiques et hédoniques. De la même manière, le score SUS est fortement corrélé avec les qualités de l'expérience utilisateur. Ces résultats présentent des pistes de recherche intéressantes, notamment pour la conception du jeu, en particulier concernant les interactions et les modules pédagogiques à intégrer.

4.4 Limitations et Perspectives

La limitation la plus importante concerne la taille relativement faible du jeu de données étudié. Ainsi, les résultats discutés doivent être renforcés par des études supplémentaires impliquant un plus grand nombre de participants. D'autre part, cette taille limitée ne permet pas d'effectuer des analyses statistiques très approfondies. Toutefois, nous espérons recueillir davantage d'informations au cours des phases expérimentales en 2025 dans le cadre du projet Agr-eSchool. Ces données additionnelles devraient nous permettre d'effectuer des analyses mettant en lumière les liens entre les variables étudiées dans le présent article et des informations dont nous ne disposons pas actuellement, telles que les interactions dans le jeu sérieux. Ces analyses pourraient nous aider à mieux comprendre le comportement des utilisateurs et ainsi nous permettre de proposer des interactions avec le plateau de jeu et entre co-équipiers afin d'améliorer leur expérience utilisateur et/ou leur sentiment d'auto-efficacité dans le contexte d'enseignement. En calculant les effets des interactions et des modules pédagogiques, nous pourrions compléter les modèles théoriques existants de conception de jeux sérieux. Une autre limitation vient de l'absence actuelle de données portant sur le gain d'apprentissage. Cette dimension est bien sûr très intéressante, mais elle n'a pas fait l'objet de notre étude afin de ne pas surcharger les enseignants avec des tâches supplémentaires. En revanche, cette partie sera traitée dans nos futurs travaux. À court terme, nous envisageons de dispenser des questionnaires pré- et post-expérimentation pour mesurer les gains d'apprentissage relatifs aux concepts étudiés au travers du jeu. À plus long terme, nous pourrons les mesurer si les mêmes élèves participent à plusieurs expérimentations, ce que nous ne sommes pas en mesure de garantir à ce jour.

5 Conclusions

Nous avons répondu à la **QR1 : L'introduction d'un jeu sérieux dans un cursus professionnel a-t-elle un impact sur le sentiment d'auto-efficacité des apprenants ? Si oui, quels sont les profils d'apprenants pour lesquels l'impact est le plus important lorsque l'on considère leur motivation et la fréquence avec laquelle ils jouent à des jeux vidéos ?** en vérifiant qu'il y a bien une différence positive significative du sentiment d'auto-efficacité après que les élèves aient joué au jeu sérieux. De plus, nous avons exploré les tendances des élèves qui ont ressenti le plus et le moins de variation d'auto-efficacité. En outre, nous avons répondu à la **QR2 : Quelle est**

l'utilisabilité et la qualité de l'expérience utilisateur perçues du jeu sérieux ? en mesurant l'utilisabilité et la qualité de l'expérience utilisateur du jeu sérieux. L'étude de l'utilisabilité décrit le jeu comme utilisable et acceptable (en comparaison avec d'autres applications web). De plus, nous avons trouvé des moyennes positives dans les qualités de l'expérience utilisateur. Finalement, nous avons exploré les relations de corrélation entre les variables précédemment citées en adressant la **QR3 : L'utilisabilité et l'expérience utilisateur perçues du jeux sérieux sont-elles corrélées avec l'auto-efficacité et la motivation des apprenants ?** À ce moment, nous n'avons pas trouve des corrélations entre la variation de l'auto-efficacité et les autres variables. Cependant, nous avons constate qu'il y a bien de corrélations significatives entre les dimensions de l'expérience utilisateur et le score SUS. Nous espérons que cette étude puisse encourager les chercheurs à explorer d'avantage l'application des jeux sérieux et analyser ses effets sur les élèves, avec le but final de contribuer à la transformation de l'enseignement agricole et d'améliorer son attractivité.

Remerciements

Ce projet est retenu dans le cadre de Deffinum, soutenu par France 2030 et le programme de l'Union Européenne NextGenerationEu. Les auteurs remercient également les partenaires du projet : HyperFiction, Succubus et le réseau UNREP.

References

1. Aaron Bangor, P.T.K., Miller, J.T.: An empirical evaluation of the system usability scale. International Journal of Human–Computer Interaction **24**(6), 574–594 (2008). https://doi.org/10.1080/10447310802205776
2. Abt, C.C.: Serious Games. The Vicking Press (1970)
3. Alvarez, J.: Du jeu vidéo au Serious Game. Theses, Université Toulouse (Dec 2007), https://hal.science/tel-01240683
4. Alvarez, J., Djaouti, D.: An introduction to serious game definitions and concepts. In Proceedings of the Serious Games & Simulation for Risks Management Workshop pp. 11–15 (01 2011)
5. Bandura, A., Carré, P., Lecomte, J.: Auto-efficacité: Comment le sentiment d'efficacité personnelle influence notre qualité de vie. Ouvertures Psychologiques, de Boeck supérieur (2019), https://books.google.fr/books?id=PYC3DwAAQBAJ
6. Bandura, A.: Self-efficacy: toward a unifying theory of behavioral change. Psychological review **84**(2), 191 (1977), publisher: American Psychological Association
7. Bangor, A., Kortum, P., Miller, J.: Determining what individual sus scores mean: Adding an adjective rating scale. Journal of usability studies **4**(3), 114–123 (2009)
8. Bardon, E., Gozard, D., Ricard, B.: L'orientation des élèves vers l'enseignement agricole et son attractivité. Tech. rep., Conseil Général de l'alimentation, de l'agriculture et des espaces ruraux (2020), https://agriculture.gouv.fr/lorientation-des-eleves-vers-lenseignement-agricole-et-son-attractivite-0

9. Benet Rivière, J.: L'empreinte du handicap dans l'orientation dans l'enseignement professionnel agricole. In: 10e congrès de l'Association française de Sociologie (AFS) " Intersections, circulations ",. AFS, Lyon, France (Jul 2023)
10. Bergeron, B.: Developing serious games. Charles river media (2006)
11. Boyer, J.M., Delattre, N.: Enseignement agricole : l'urgence d'une transition agro-politique. Tech. rep., Sénat (2021), https://www.senat.fr/notice-rapport/2020/r20-874-notice.html
12. Brooke, J.: Sus: A quick and dirty usability scale. Usability Eval. Ind. **189** (11 1995)
13. Djaouti, D., Alvarez, J., Jessel, J.P.: Classifying serious games: the g/p/s model. Handbook of Research on Improving Learning and Motivation through Educational Games: Multidisciplinary Approaches (01 2011). https://doi.org/10.4018/978-1-60960-495-0.ch006
14. Espinosa-Curiel, I.E., de Alba-Chávez, C.A.G.: Serious video games for agricultural learning: scoping review. IEEE Transactions on Learning Technologies (2024)
15. FRANCE2030, F.: Agr-eschool (2023-2025), https://agreschool.fr/
16. Gómez-Prada, U., Orellana-Hernández, M., Salinas-Ibáñez, J.: Strategy for the appropriation of a DSS in small bovine producers using simulation and a serious video game. Information **11**(12), 566 (2020), publisher: MDPI
17. Gómez Prada, U.E., Orellana Hernández, M.L., Salinas Ibañez, J.M.: Systems dynamics and serious video games in an appropriation strategy of a decision support system for small livestock farmers. International Journal of Interactive Mobile Technologies (iJIM) **14**(15), pp. 4–24 (Sep 2020). https://doi.org/10.3991/ijim.v14i15.14597, https://online-journals.org/index.php/i-jim/article/view/14597
18. Hassenzahl, M.: Experience design: Technology for all the rightreasons. Synthesis Lectures on Human-Centered Informatics (2010), http://dx.doi.org/10.2200/S00261ED1V01Y201003HCI008
19. Hassenzahl, M., Burmester, M., Koller, F.: AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität, pp. 187–196. Vieweg+Teubner Verlag, Wiesbaden (2003). https://doi.org/10.1007/978-3-322-80058-9_19
20. Hassenzahl, M., Tractinsky, N.: User experience - a research agenda. Behaviour and Information Technology **25**, 91 – 97 (03 2006). https://doi.org/10.1080/01449290500330331
21. ISO: Iso 9241-11:2018. ergonomie de l'interaction homme-système. Tech. rep., ISO (2019)
22. Lallemand, C., Koenig, V., Gronier, G., Martin, R.: Création et validation d'une version française du questionnaire attrakdiff pour l'évaluation de l'expérience utilisateur des systèmes interactifs. European Review of Applied Psychology **65**(5), 239–252 (2015). https://doi.org/10.1016/j.erap.2015.08.002
23. Lecomte, J.: Les applications du sentiment d'efficacité personnelle. Savoirs, Hors série pp. 59–90 (2004). https://doi.org/0.3917/savo.hs01.0059.
24. OECD/FAO: Agricultural outlook 2022-2031 (2022). https://doi.org/10.1787/f1b0b29c-en
25. Pelletier, C., Gaudreau, N., Frenette, É.: Portrait des pratiques enseignantes qui soutiennent le sentiment d'efficacité personnelle à réussir ses études au collégial selon les perceptions étudiantes. Revue des sciences de l'éducation **48**(2) (2022). https://doi.org/10.7202/1098345ar

26. Pintrich, P., Smith, D., Duncan, T., Mckeachie, W.: A manual for the use of the motivated strategies for learning questionnaire (mslq). Ann Arbor. Michigan **48109**, 1259 (01 1991)
27. Pons Lelardeux, C.: Real-time Virtual Collaborative Environment Designed for Risk Management Training: Communication and Decision Making. Theses, Université Paul Sabatier (Toulouse 3) (Oct 2017), https://hal.science/tel-02008969
28. Sauro, J.: A Practical Guide to the System Usability Scale: Background, Benchmarks & Best Practices. Measuring Usability LLC (2011), https://books.google.fr/books?id=BL0kKQEACAAJ
29. Sauro, J., Lewis, J.R.: Chapter 8 - standardized usability questionnaires. In: Sauro, J., Lewis, J.R. (eds.) Quantifying the User Experience (Second Edition), pp. 185–248. Morgan Kaufmann, Boston, second edition edn. (2016). https://doi.org/10.1016/B978-0-12-802308-2.00008-4
30. Tangworakitthaworn, P., Tengchaisri, V., Sudjaidee, P.: Serious game enhanced learning for agricultural engineering education: Two games development based on iot technology. In: 2020-5th International Conference on Information Technology (InCIT). pp. 82–86. IEEE (2020)
31. Vallerand, R., BLAIS, M., Brière, N., Pelletier, L.: Construction et validation de l'Échelle de motivation en Éducation (eme). Canadian Journal of Behavioural Science-revue Canadienne Des Sciences Du Comportement **21**, 323–349 (07 1989). https://doi.org/10.1037/h0079855
32. Whyte, M., Smyth, M., Scherf, K.: Designing serious game interventions for individuals with autism. Journal of Autism and Developmental Disorders **45**, 3820–3831 (2015)

La Pertinence du Catalogue du Compte Personnel de Formation, vue au Prisme du Comportement des Utilisateurs

Matthieu Cisel

CY Cergy Paris Université, Laboratoire AGORA, Cergy, France
matthieu.cisel@cyu.fr

Résumé. En France, le Compte Personnel de Formation (CPF) sert de véhicule aux politiques publiques visant à financer la formation d'adultes. La plateforme Mon Compte Formation (MCF) permet de sélectionner la formation de son choix au sein d'un catalogue particulièrement fourni; elle met à disposition un certain nombre de données sur son offre, données qui à ce jour restent sous-exploitées. Nous présentons ici une réflexion méthodologique sur les opportunités offertes par ces traces. Elle vise à apporter un regard sociologique et géographique sur des analyses de traces réalisées par des méthodes issues de l'informatique, pour mieux contraster l'offre des catalogues de formation et les thématiques qui constituent les principaux centres d'intérêt du public de la plateforme, ceux-ci étant inférés sur la base des traces laissées via le moteur de recherche intégré au sein de MCF.

Mots-clés : CPF, fouille de règles d'association, formation d'adultes, traces, organisme de formation

Abstract. In France, the Personal Training Account (CPF in French) serves as the main tool for public policies aimed at financing adult training. The Mon Compte Formation (MCF) platform allows you to select the training of your choice from a particularly comprehensive catalog; it makes available a certain amount of data on its offer, data which to date remains under-exploited. In particular, we present analyzes mobilizing geographical approaches, that aim to highlight the interest in methods from computer science to better understand catalog construction strategies, as well as the possible inadequacy of the catalogs proposed in relation to the expectations of the public, expectations that can be analyzed thanks to the logs left in research engines used within MCF.

Keywords: CPF, association rule mining, lifelong learning, analytics, training establishment

1. Introduction

1.1 Contexte de l'étude et questions de recherche

En matière de formation d'adultes, le Compte Personnel de Formation (CPF) représente en France un instrument central des politiques publiques [18]. Rattaché à un individu plutôt qu'à son contrat de travail, il vise à offrir des droits continus à la formation tout au long de la vie. Il assure ainsi un accès ininterrompu à des droits [7], et permet aux salariés une grande liberté dans le choix de leurs formations [16,19]. Ainsi, il n'est pas nécessaire de choisir une formation en fonction de la profession exercée, ni de demander l'approbation de l'employeur. La plateforme qui héberge le catalogue de formations se nomme Mon Compte Formation (MCF) depuis 2019 ; elle est gérée par la Caisse des Dépôts et Consignations (CDC).

Le lancement de la plateforme MCF constitue l'un des nombreux avatars de ce qu'un auteur comme Bullich [9, 10] nomme la « plateformisation » de la formation. Dimension centrale de cette plateformisation, des milliers d'organismes de formation (OF), en concurrence les uns avec les autres, proposent sur MCF une offre hétéroclite, allant de l'informatique au permis de conduire en passant par toutes sortes de titres et certifications. Les chercheurs d'emploi constituent un contingent important de la plateforme, mais des publics variés sont représentés [5]. La CDC s'intéresse de près aux retours sur investissement associés à la mise en place du CPF, en contribuant à des études sur les choix des formations achetées selon les profils d'apprenants, mais aussi en analysant les traces issues de la navigation sur MCF. Le premier type d'analyse fait l'objet de publications ouvertes au grand public [6,13,15,17]. Le second relève en revanche de réflexions menées uniquement en interne, conséquence des contraintes, politiques ou légales, associées à la diffusion de travaux mobilisant des traces [4].

Pourtant, ces dernières présentent un intérêt certain tant du point de vue des visées de la recherche que de celles, plus pragmatiques, des organismes de formation souhaitant disposer d'informations pertinentes pour développer leur catalogue de cours. Comprendre les déterminants de ce phénomène est crucial en termes d'efficacité de l'action politique, qui peut orienter les financements. Des entreprises comme Wedof[1] proposent toutes sortes de services aux OF, mais faute d'accès aux traces, détenues par la CDC, ils ne peuvent se situer au niveau de granularité que nous proposons ici. Notre démarche pourrait donc aussi bien informer des politiques publiques que servir de base à la conception de services à destination des organismes de formation.

Parmi les nombreuses opportunités qu'offre l'exploitation des traces, nous nous proposons dans cet article de discuter de la question de l'adéquation entre offre proposée par les OF et centres d'intérêt des utilisateurs de MCF, ceux-ci étant inférés sur la base de leur comportement de navigation sur la plateforme. Le choix du terme à employer a été l'objet de discussions : si la visée initiale du projet était de contraster offre et attentes, nous ne pouvons pas inférer quant aux attentes sur la base du seul comportement de navigation. Le fait est que les attentes d'un individu peuvent difficilement être capturées par de simples comportements ; une enquête s'impose. Le terme « centre

[1] https://www.wedof.fr/

d'intérêt » est plus neutre qu'attente ou demande, puisqu'il est plus légitime de supposer qu'un apprenant porte un certain intérêt à une thématique ou une formation dont il lit le descriptif ou qu'il entre dans un moteur de recherche.

En matière d'évaluation des politiques publiques, les enjeux d'études sur la pertinence des catalogues de formation sont considérables. Chaque année, l'Etat, ses opérateurs, les OPCO (opérateurs de compétences), et les entreprises investissent plus de trois milliards d'euros dans le CPF. Par ailleurs, cette somme ne prend pas en compte le fait qu'une partie significative des apprenants financent par eux-mêmes une part des frais d'inscriptions, à plus forte raison depuis le vote de l'amendement n°698 de l'article 49 du Code du Travail, en 2023 [11]. La bonne utilisation de ces sommes dépend largement de la pertinence du catalogue pour l'audience de la plateforme. Cette considération nous a amené à structurer le présent article autour de questions de recherche qui fait de cette contribution une réflexion de nature avant tout méthodologique structurée autour de deux questions :

Comment les données disponibles tant sur le catalogue de formation de MCF que sur le comportement de navigation au sein de la plateforme, et du moteur de recherche afférent, peuvent-elles être combinées pour évaluer l'adéquation de l'offre avec les thématiques de formation suscitant l'intérêt des apprenants ? Dans quelle mesure les données à disposition de la CDC permettent-elles d'analyser sous l'angle géographique et sociologique une telle adéquation ?

Aborder ces questions implique en premier lieu de construire des indicateurs sur la base des traces résultant de recherches, ou de la navigation au sein de MCF. Cette contribution mêle des résultats obtenus sur la version de juin 2023 du catalogue de MCF, pour ce qui est de l'offre, et de données simulées, pour ce qui est des traces. En effet, si des analyses de traces réelles ont été produites par nos étudiants pour le compte la CDC dans le cadre de leurs stages, les résultats correspondants ne peuvent être publiés dans cette contribution, du fait des contraintes politiques et légales [4]. Néanmoins, dans la mesure où il s'agit d'une preuve de concept, la substitution de dernière minute par des données artificielles ne change rien à la visée du propos.

1.2 De l'approche géographique des traces à la notion de visite

Nous avons mené la présente étude en premier lieu sous l'angle des thématiques recherchées par l'utilisateur, que l'on peut ensuite contraster avec les thématiques proposées dans les catalogues, mais aussi sous l'angle des autres caractéristiques des formations en question, comme le niveau d'études correspondant. Le croisement entre d'une part des données sociologiques et économiques récoltées à l'échelon national et à une granularité fine – celle de la commune, et d'autre part des traces issues de plateformes, a été réalisé dans plusieurs études américaines [12, 20], mais n'est pas une pratique commune dans les recherches en France.

S'agissant plus précisément de l'interaction de cohortes d'apprenants avec une offre de cours, nous avions déjà illustré sur la base d'analyses de traces d'une plateforme de MOOC comment des centaines de milliers d'apprenants voyaient leur comportement de sélection de formation influencé par les caractéristiques d'un catalogue [2,3].

Cette recherche a employé plusieurs concepts, notamment celui de "clé d'entrée", qui désigne le processus ayant mené à la sélection d'un produit, correspondant ici à l'inscription à un cours. Ce concept se divise en deux catégories : la "clé d'entrée plateforme" et la "clé d'entrée produit". D'une part, un internaute peut visiter un site en ayant une idée générale de ce qu'il veut acheter, et d'autre part, il peut vouloir acquérir un ou plusieurs produits sans connaître précisément les sites où les trouver. Dans le premier cas, la clé d'entrée est le site marchand à consulter, tandis que dans le second, il s'agit du produit spécifique à acheter [13].

Dans le contexte de la « visite expérientielle », les motivations sont principalement hédonistes, incluant la découverte d'offres originales, le plaisir de la surprise, et la promenade à travers un vaste assortiment de produits, évoquant la logique du lèche-vitrine. Des visites régulières sur un site de e-commerce sont des exemples de ce type d'achat expérientiel [13, p. 213]. Cela s'applique également aux participants qui, après avoir découvert MCF, y retournent fréquemment pour voir si de nouveaux cours susceptibles de les intéresser ont été ajoutés, même s'ils n'ont pas d'idée précise sur le thème de formation qu'ils recherchent. Dans le cas de la « visite expéditive », « l'internaute se rend sur un site pour réaliser un achat assez précis. Soit il connaît l'offre de ce site pour ce produit (il a vu une offre sur une publicité ou un ami lui en a parlé, par exemple) ; soit il sait […] que ce site est le meilleur pour telle catégorie de produit. […] À la différence du précédent, ce processus de décision est guidé par un objectif précis : les motivations de visite sont plutôt utilitaires » [13, p. 214].

2. Méthodologie

La réflexion sur le contraste entre offre et demande a été menée en parallèle d'une part au sein du laboratoire AGORA, et d'autre part par des élèves-ingénieurs de CY Tech effectuant dans le cadre de leur stage de fin d'études des missions d'analyse de traces pour le compte de la CDC. Les informations sur les traces rapportées ici sont issues de leur mémoire de fin d'étude. Ne pouvant rapporter les graphiques produits dans ce cadre, contraintes légales obligent, nous avons présenté les données collectées quotidiennement via MCF, puis généré des données artificielles pour illustrer notre propos. Nous revenons ensuite sur l'approche suivie en parallèle pour étudier les catalogues de formation, dont les données sont mises à disposition du grand public.

2.1 Des traces de navigation collectées notamment par Matomo

La CDC dispose d'une table contenant spécifiquement les logs sur les requêtes effectuées dans le moteur de recherche interne à MCF. En plus de l'horodatage, ils incluent la thématique recherchée (enregistrée via son formacode), la zone géographique (pour les formations non dispensées à distance), d'éventuels filtres mobilisés : certains sont obligatoires à chaque recherche, comme la modalité (face-à-face/présentiel), d'autres non. L'on peut par exemple effectuer des recherches par organisme de formation.

Les indicateurs relatifs aux recherches sont utiles pour inférer quant aux centres d'intérêt du public de la plateforme, mais ils ne représentent qu'un aspect de la question.

L'on pourrait par exemple afficher en complément, par zone géographique, le nombre de dossiers refusés faute de place dans les formations. L'on pourrait ainsi mieux identifier les thématiques pour lesquelles la demande dépasse l'offre, mais l'on sortirait ce faisant du champ des learning analytics, et donc des EIAH. Or l'objectif de cette contribution est justement de fournir une illustration supplémentaire de l'intérêt qu'il y a à se placer à l'intersection entre sciences humaines et sociales et learning analytics.

En plus des traces produites par le moteur de recherche, la CDC dispose de tables comprenant des données détaillées sur les événements produits par les interactions avec le site : visiter une page, cliquer sur un lien externe ou sur un téléchargement, recherche sur le site. Si toutes les actions sont stockées dans les serveurs de la CDC, le trafic est en partie suivi grâce à *Matomo*, outil grand public régulièrement mobilisé pour analyser les utilisations d'un site Internet, mais que nous ne mobilisons pas ici. Il permet une sélection des événements d'intérêt fournit une interface d'interaction ainsi qu'une API dont la fonction est d'extraire des traces précises par requêtage. Dans certaines conditions, il est possible de relier le *user_id* stocké par Matomo avec les données dites du *Dossier*, qui sont des informations autodéclarées fournies par l'apprenant : dernier diplôme obtenu, âge, genre, statut socio-professionnelle. Le numéro de sécurité sociale, identifiant utilisé dans MCF, est exclu de ces données.

Nous avons produit, pour mimer les requêtes dans le moteur de recherche, des traces artificielles ayant strictement la même composition que les données réelles permet de contourner les obstacles légaux liés à la diffusion par la CDC de traces réelles, sans pour autant devoir renoncer à illustrer les concepts sous-tendant les indicateurs. Des étudiants formés par nos soins ont fourni dans leurs rapports, suite à un stage portant sur l'analyse des données brutes de la CDC, des illustrations de la structure des bases de données en jeu, avec des exemples permettant de montrer une trace typique, pour une lecture de descriptif de formation, ou pour une requête. Ces éléments sont suffisants pour produire des bases identiques, sur le plan des variables en jeu, à celles manipulées par les analystes de la Caisse des Dépôts.

Les cartes que nous produisons dans la section dédiée aux résultats ne semblent pas constituer un risque majeur d'exposition de données à caractère privé, et nous souhaitons dès lors nous attarder ici sur les raisons qui nous ont poussé à produire des données artificielles. En effet, il faut pour les produire passer par des étapes où l'on est capable d'identifier les termes entrés par un individu donné dans le moteur de recherche de MCF ; se posent donc des questions de données à caractère personnel, et de conventions lourdes à mettre en place.

La CDC est soumise au régime du RGPD, comme, au demeurant, tout opérateur de l'Etat. Si elle est légitimement en droit de mener des analyses sur les traces que MCF produit, elle ne peut pas transmettre ces données à des laboratoires de recherche sans la signature de conventions appropriées. De fait, ce n'est à notre connaissance qu'en 2023 que les conditions générales d'utilisation (CGU) de MCF ont inclus la possibilité d'utiliser les données selon une logique de recherche, et la signature des CGU mises à jour a donc constitué un préalable pour qu'un individu puisse être inclus dans une étude. Le travail de développement d'une relation professionnelle de confiance entre acteurs de la recherche et opérateurs de l'Etat, le processus de conventionnement, généralement long, l'exclusion, hors des bases manipulées, des individus n'ayant pas signé les CGU

mises à jour : toutes ces étapes, souvent laborieuses, n'ont de sens que si l'étude porte sur des phénomènes capturés dans les traces. Elles sont impératives si la focale, par exemple, porte sur l'évolution de la nature des thématiques les plus plébiscitées dans tel ou tel département français. Dans la mesure où nous ne cherchons pas à nous inscrire dans la mise en évidence d'un phénomène, mais plutôt à apporter un regard méthodologique sur la production d'indicateurs d'une part, et à mettre en lumière les opportunités offertes par les bases de MCF d'autre part, il ne nous a pas semblé pertinent d'investir à ce stade dans la mise en place d'une collaboration formelle avec l'opérateur de l'Etat.

2.2 Analyse de l'offre de formation

La CDC met à disposition un jeu de données nommé *offre de formation*, mis à jour régulièrement sur son site[2]. Les analyses présentées ici correspondent à la version du 1er juin 2023. Nous avons analysé en premier lieu la base sous l'angle de la géographie, l'échelle retenue étant celle du département. En effet, à moins que l'utilisateur ne cherche à se former à distance, il recherche a priori le cours, en priorité dans son environnement immédiat, c'est-à-dire sans son département, ou éventuellement dans un département limitrophe à son lieu de résidence. Précisons que l'on ne peut connaître l'adresse de résidence d'un utilisateur connecté à la plateforme MCF que s'il s'est connecté – il est en effet possible de naviguer sans se connecter. C'est donc à la richesse de l'offre dans l'environnement géographique proche qu'il est intéressant de contraster le résultat des recherches effectuées – à moins que l'on se penche sur les formations à distance.

S'agissant de la variable retenue pour représenter les centres d'intérêt, nous avons priorisé la thématique et le niveau de sortie, équivalent à un niveau d'études obtenu à la suite de la formation. La thématique peut être vue à plusieurs échelles, celle du *Formacode* (plusieurs milliers), celle du *Champ sémantique* (aussi appelé *NSF*, au nombre de 65, comme *Pêche, aquaculture*), et celle du *Grand domaine* (au nombre de 14, comme *Agriculture, Environnement*). Nous avons par exemple travaillé avec la version du printemps 2023 de la taxonomie, disponible via le Centre Inffo[3]. Elle comprend un identifiant numérique, le libellé (Boulangerie, etc.), le grand domaine correspondant, sa place dans l'arborescence des formacodes.

[2] https://opendata.caissedesdepots.fr/
[3] https://formacode.centre-inffo.fr/

2.3 Synthèse des indicateurs présentés

Sur la base des données relatives au moteur de recherche interne, de consultations de descriptifs et d'analyses de catalogue, de nombreux indicateurs peuvent être construits. Nous en proposons un récapitulatif dans le Tableau 1.

Indicateur	Mode de construction et limites
% de recherches positives	Analyse de la nature des requêtes issues de l'utilisation du moteur de recherche de MCF : mots-clés sélectionnés ou entrés dans le champ de texte libre. Si le moteur présente dans sa réponse au moins un élément correspondant effectivement à ces mots-clés, la recherche est alors considérée comme « positive ».
% d'une thématique donnée dans un catalogue de formation	100 * le nombre des formations correspondant à une thématique donnée au sein d'une aire géographique divisé par le nombre de formations proposées dans cette aire. Mobilisation de la base en open data (données réelles) nommée *offre de formation*.
Ratio demande / offre	% de la thématique dans les requêtes, divisé par le % de la thématique, en nombre de formation, au sein de l'offre locale

3. Résultats

Nous présentons en premier lieu des analyses mêlant traces artificielles et analyse du catalogue réel pour contraster offre et demande, pour discuter une série de métriques jugées pertinentes pour mesurer l'adéquation entre les deux.

3.1 Analyses des données du moteur de recherche

Comme les recherches réalisées dans MCF requièrent systématiquement d'entrer un code postal lorsque l'on recherche une formation en présentiel – en précisant une distance par rapport à la commune recherchée, il est possible de réaliser des cartes à l'échelle de la commune. Ainsi, plusieurs indicateurs peuvent être construits et représentés sous la forme de cartes. Nous représentons ci-dessous des données simulées analogues aux données réelles, dans lesquels nous représentons un indicateur d'adéquation, en réalisant l'analyse avec d'une part avec tous les utilisateurs de MCF (Figure 1, Haut), et d'autre part ceux ayant déclaré n'avoir que le Bac (Figure 1, Bas). Dans ce dernier cas, l'objectif est d'introduire une dimension sociologique en plus de l'aspect géographique.

En Fig. 1, nous avons représenté le taux de recherches positives, c'est-à-dire le pourcentage des requêtes ayant retourné au moins un résultat pour la thématique recherchée, pour le niveau de distance à la commune sélectionné (qui peut être choisi par

l'utilisateur entre 5, 10, 15, 50 ou 100 km par exemple), par contraste avec les recherches infructueuses. Il nous faut faire le distinguo ici entre une recherche, qui nous donne des indications sur une intention, et une consultation, qui implique qu'un résultat a été trouvé, et que le descriptif de la formation a été visionné.

Nous avons retenu ici un rayon de 5 km autour de la commune sélectionnée, tous grands domaines confondus ; ce seuil peut être modifié de manière arbitraire, selon que l'on s'intéresse aux formations situées dans l'environnement immédiat de l'apprenant, ou s'il lui est possible de se déplacer jusqu'à une certaine distance. On remarque pour ce jeu artificiel, dans le premier cas (Gauche), des valeurs plus élevées pour des centres urbains tels que Quimper, ou Rennes, signe que les habitants de ces communes trouvent davantage ce qu'ils cherchent via le moteur de MCF. Il suggère alors une meilleure adéquation de l'offre aux apprenants vivant en communes urbaines par rapport aux communes rurales, peut alors être utilisé dans la conception de politiques publiques visant à atténuer les inégalités d'accès à la formation entre la ville et la campagne, notamment via l'encouragement à la production de formations à distance. L'on peut également aborder la question sous un angle sociologique, en contrastant par exemple le pourcentage de recherches positives selon que le plus haut diplôme obtenu soit le baccalauréat, le master ou le doctorat.

En filtrant par le niveau de formation des apprenants (Figure 1, Bas), accessible par le dossier, l'on constate que les personnes peu diplômées sont logées à la même enseigne que les autres dans leurs recherches, sans distinction majeure entre les grandes villes et les communes rurales. L'interprétation de tels résultats nécessite ensuite un travail statistique, fondé sur des régressions, visant à identifier les variables sociodémographiques et socioprofessionnelles associées susceptibles d'expliquer des inégalités géographiques.

Comme nous l'avons souligné dans la section consacré à la méthodologie, il convient de compléter les analyses de traces du moteur de MCF par des analyses de consultations de descriptifs. En effet, un certain nombre d'apprenants découvrent ces descriptifs par des recherches sur des moteurs classiques comme Google, et non par le moteur interne à MCF. Par contraste avec le cas précédent, l'on ne peut ici parler de recherche infructueuse, puisque par définition une formation a été trouvée et son descriptif consulté.

Aborder la question sous un angle géographique est moins intuitif, puisque la formation est certes associée à une adresse, mais il n'est pas obligatoire pour l'apprenant de spécifier sa propre adresse pour atteindre et visionner le descriptif. Une recherche sur Google ne permettant pas de prendre en considération des critères de distance au domicile, elle conduit a priori à découvrir fréquemment des formations organisées trop loin du lieu de résidence de l'apprenant, ce qui vient interférer avec une analyse sur la bonne couverture géographique des besoins de formation.

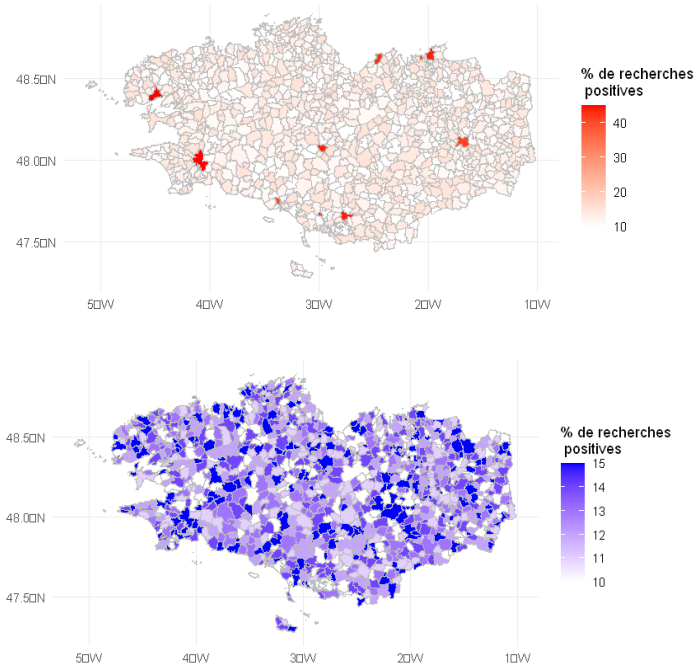


Fig. 1. Carte représentant, pour les communes de Bretagne, le pourcentage de recherches positives, tous utilisateurs confondus (Haut), et utilisateurs dont le diplôme maximal obtenu est le bac (Bac). Données simulées.

3.2 Contraster offre et demande : un regard géographique et sociologique

Nous revenons maintenant sur les indicateurs qui ne peuvent être construits seulement sur la base de traces, mais qui impliquent une mise en regard avec l'offre. Sur le plan géographique, il n'est pas pertinent de représenter à l'échelle de la commune, car il faut un nombre suffisamment important d'OF par unité géographique pour produire des statistiques pertinentes, ce qui invite à ne pas descendre en dessous du niveau départemental. Il s'agit, en se concentrant sur les formations dispensées présentiel, de calculer le ratio entre le pourcentage des apprenants effectuant une recherche dans une thématique donnée (niveau formacode, NSF ou Grand domaine), et le pourcentage des formations de l'unité géographique (du département par exemple) correspondant à cette thématique. S'il faudrait proposer des indicateurs plus fins prenant en compte le nombre de places de formation ouvertes, et non simplement le nombre de formations disponibles, l'on peut néanmoins voir que plus cet indicateur élevé, plus la demande est importante par rapport à l'offre, et inversement.

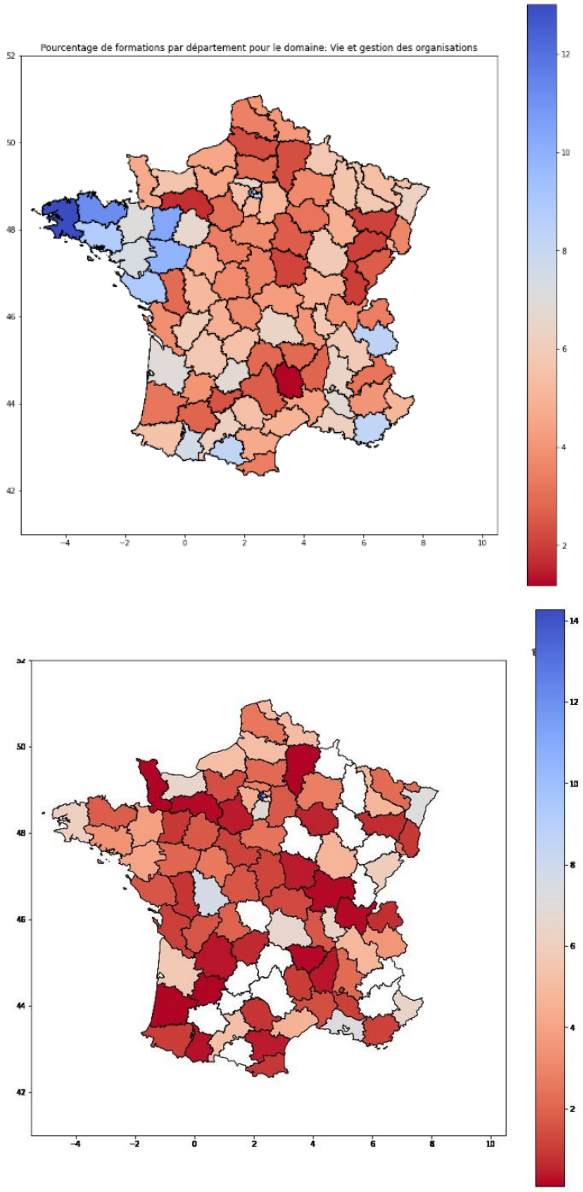


Fig. 2. Carte du pourcentage des formations du catalogue de MCF daté de juin 2023 relevant du domaine : Vie et gestion des Organisations (Haut) et pour les formations de niveau 7 (équivalent Master), les valeurs les plus élevées étant Paris (Bas). Données réelles.

En prenant en compte le catalogue de formation, et non plus seulement les traces laissées par les apprenants putatifs, nous gagnons en profondeur d'analyse. Sur la base des données fournies par la CDC, nous avons observé des disparités géographiques conséquentes eu égard à la part d'une thématique dans l'offre de formation. A des fins d'illustration, la Figure 2 (Haut) présente la place qu'occupe le grande domaine « Vie et gestion des organisations » dans le pourcentage de l'offre de formation d'un département ; cette thématique est par exemple surreprésentée en Bretagne. Les OF se positionnent largement sur ce sujet, sans que l'on ne sache précisément si la démarche a son équivalent dans les demandes.

Dans la Figure 2 (Bas), nous abordons la question de l'offre sous l'angle du niveau de sortie, représentant l'un des niveaux les plus élevé : le niveau 7, qui est un équivalent Master. De nombreux départements ne proposent qu'une offre limitée dans ce domaine, voire aucune offre – ils sont alors représentés en blanc. Seul Paris, avec 14% de l'offre, fait figure d'exception, sans doute car l'on y trouve un nombre important d'individus fortement diplômés, dont le taux de recours au CPF est également largement supérieur à la moyenne nationale [1], ce qui, a priori, stimule la création d'une offre distincte de celle du reste du territoire.

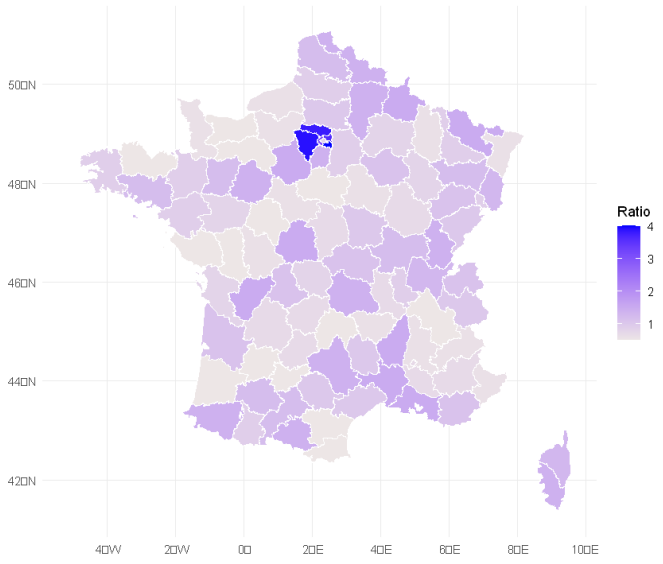


Fig. 3. Carte du ratio demande / offre, mesuré sur la base des requêtes dans le moteur de recherche interne, illustration pour la thématique Sports, Loisir, Tourisme. Données simulées.

Pour établir le lien entre offre et demande, nous proposons en Figure 3 une carte présentant un ratio entre le % de la thématique dans les requêtes, divisé par le % de la thématique, en nombre de formation, au sein de l'offre locale. En Île de-France, on constate pour le Grand domaine *Sports, Loisir, Tourisme* le rapport entre la demande, telle que perçue dans les requêtes du moteur de MCF, est nettement plus prononcée par rapport à l'offre existante que dans les autres régions de France. Une telle information pourrait être utile pour des OF qui envisageraient de développer ou d'enrichir leur offre dans ce secteur, en leur fournissant des indications quant aux dynamiques l'œuvre dans cette zone géographique.

4. Discussion

Alors que les travaux de la Direction de l'Animation de la Recherche, des Études et des Statistiques (DARES) et de la CDC montrent que les analystes en charge de rendre compte des utilisations de MCF sont en capacité de traiter sous l'angle des SHS des données de consommation du CPF, et ce pour le compte de l'Etat, les ponts avec l'informatique restent à construire. Les lacunes en la matière sont manifestes dans la littérature scientifique. Comme nous l'avons souligné en introduction, malgré l'existence de réflexions méthodologiques sur les traces [19] et de cas d'études ancrés empiriquement [12], le nombre de publications sur le sujet reste faible. C'est la raison pour laquelle nous avons proposé ce travail sur la conception d'indicateurs permettant d'aller au-delà des analyses de données d'achat de formation faites par la DARES, qui ne couvrent qu'un seul des aspects du comportement d'achat et de navigation des utilisateurs de la plateforme.

En effet, ces achats peuvent être faits « faute de mieux », pour dépenser le solde disponible dans les OF présents dans l'environnement immédiat, ou sous l'influence d'organismes peu scrupuleux réalisant des campagnes d'appels téléphoniques pour faire la promotion de leur offre. En mobilisant l'informatique, et plus précisément l'analyse de traces laissées via les moteurs de recherche, l'on peut avoir un accès plus direct à la nature des formations recherchées par les apprenants, et identifier si le catalogue contient des éléments correspondant aux recherches effectuées. Néanmoins, tout indicateur présente ses limites et nous nous proposons de mobiliser les notions de visites présentées en introduction pour suggérer des alternatives [2], en mobilisant celles de visite expérientielle et expéditive. Précisons notre propos, qui porte sur l'indicateur présenté en Figure 1, et toutes ses variantes. Ceci tient au fait que l'indicateur présenté en Figure 3, contrastant offre et demande, ne nécessite pas de prendre en compte cette notion de visite, puisqu'il est construit sur la base de l'agrégation d'un nombre important de requêtes.

L'indicateur centré sur les recherches positives porte le flanc à une critique en simplification excessive en ce sens qu'il met toutes les requêtes sur le même plan, aussi bien les premières recherches effectués au moment de la découverte de MCF, que celles réalisées après s'être approprié la plateforme, après plusieurs achats, et de nombreuses visites. Or l'on peut émettre l'hypothèse selon laquelle les premières thématiques recherchées par l'utilisateur sont celles pour lesquelles l'appétence est la plus marquée.

Dès lors, un utilisateur qui ne trouverait pas satisfaction aux premières tentatives se rabattra sans doute, pour dépenser le solde de son compte sur les formations disponibles dans son environnement immédiat, quitte à choisir des thématiques qui ne seraient pas son premier choix.

Il s'agit alors de modéliser les requêtes selon une séquence, et leur attribuer un rang dans cette séquence. On distinguerait alors les recherches infructueuses réalisées en début de séquence de celles qui ont lieu en fin de séquence, longtemps après, par exemple, avoir réalisé un premier achat. Lorsqu'une recherche est infructueuse, l'on ne devrait en effet sans doute pas même sur un même plan des sujets recherchés en priorité, relevant par exemple de visites expéditives où l'utilisateur se concentre sur un sujet et un seul, et des thématiques recherchées lors de visites expérientielles, par exemple. Dans ce dernier cas, l'utilisateur effectue dans un temps limité des requêtes sur des sujets variés, par exemple pour se former à devenir boulanger, instructeur de yoga, professeur des écoles ; la signification à attribuer à la requête est alors sensiblement différente. Dans la mesure où il s'agit de l'équivalent du lèche-vitrine, si pour un sujet donné, l'utilisateur de trouve pas satisfaction, il y a manifestement pour lui d'autres thématiques qui pourraient l'intéresser suffisamment pour qu'il investisse le solde de son compte CPF.

Terminons cette discussion par une réflexion sur l'application de tels indicateurs. L'intérêt est manifeste pour des décideurs officiant au sein d'entreprises privées (OF, etc.) et jouant un rôle dans la construction des catalogues de formation présents sur MCF. Identifier l'intérêt que porte l'audience de la plateforme pour telle ou telle thématique permet d'orienter une offre, à plus forte raison si l'on peut identifier les aires géographiques où l'offre est lacunaire par rapport à la demande, ce qui permet une meilleure granularité dans la prise de décision. S'agissant du pilotage des politiques publiques, l'application est moins évidente, car une grande agentivité est laissée aux acteurs de la formation continue dans la construction de leur offre. La puissance publique intervient au niveau des certifications, pour déterminer qui peut, via l'inscription au RS ou au RNCP, bénéficier des fonds du CPF. Le fait d'être considéré comme un métier en tension aide pour ce faire, et pour le moment, cette dénomination de « métier en tension » est davantage basée sur des enquêtes auprès des entreprises que sur d'autres considérations. Mais peut-être à terme peut-on envisager que la demande ou l'intérêt que présente l'audience de la plateforme pour telle ou telle thématique devienne un critère de décision ; les cartes et les indicateurs que nous avons produits pourraient alors prendre tout leur sens.

5. Conclusion

Au-delà des obstacles à la diffusion des résultats obtenus suite aux analyses de traces, la principale limite de ce type de recherche réside dans le fait qu'un utilisateur peut réaliser des recherches sur MCF sans se connecter, et donc sans qu'un identifiant utilisateur ne soit associé. Dans ce cas de figure, l'on ne dispose pas de données sur ses caractéristiques sociodémographiques et socioprofessionnelles. Cela constitue un obstacle inhérent à la réalisation des analyses que nous avons proposées.

En guise de perspective à cette recherche, nous suggérons de mettre en œuvre ces analyses en mobilisant des données plus précises que celles fournies dans le *Dossier*, ce qui passe par exemple par la mobilisation de la base nommée *Déclaration Sociale Nominative* (DSN) [8]. Au regard de la sensibilité des informations qu'elle porte, des accords contraignants devront être signés avec les chercheurs intéressés par de telles approches. Constituée par l'Etat, la base rassemble notamment des données sur les professions occupées, ainsi que les revenus associés, tout au long de la vie d'un individu.

En croisant traces et données de la DSN, l'on pourrait par exemple caractériser les comportements de recherche de formation avant ou après une transition professionnelle. Ainsi, si une recherche suit un changement de poste, peut-être est-ce le symptôme d'une volonté de développer ses compétences dans un secteur nouvellement rejoint. Si une recherche a précédé une transition, indépendamment d'un achat effectif, il est plus pertinent de l'interpréter comme une volonté d'anticiper une transition. L'on pourrait ainsi mieux appréhender les motivations des utilisateurs qui se connectent.

6. References

1. Albert, C. (2021). Le recours au Compte personnel de formation par les salariés du secteur privé : quelles évolutions entre 2019 et 2020 ? Questions Politiques Sociales - Les brèves n°10, septembre 2021, disponible sous : https://politiques-sociales.caissedesdepots.fr/qps-les-breves-ndeg10
2. Cisel, M. (2018). Le choix d'un MOOC, un processus influencé par l'organisation des cours en catalogues?. *STICEF, 25,* 121-150
3. Wintermute, E. H., Cisel, M., & Lindner, A. B. (2021). A survival model for course-course interactions in a Massive Open Online Course platform. *PloS one*, *16*(1), e0245718.
4. Cisel, M. (2022). De la collecte des traces d'interaction à la publication scientifique. Un passage à l'échelle semé d'embûches. *Distances et médiations des savoirs. 37*.
5. Balmat, C. (2018), Le compte personnel de formation : Une montée en charge rapide surtout au profit des personnes en recherche d'emploi, Dares résultats, n°004 ; Dares, janvier 2018. https://dares.travail-emploi.gouv.fr/IMG/pdf/2018-004.pdf
6. Bousquet, G., et Jaumont, L. (2020). Le compte personnel de formation pour les salariés : un retour sur les coûts de formation 2018, Questions retraite et solidarité, n°30, février 2020, disponible sous : https://retraitesolidarite.caissedesdepots.fr/sites/default/files/qrs_29.pdf
7. Boutinet, J.-P. (2019). Dans le champ de la formation professionnelle, quel partage de responsabilités pour aujourd'hui ? *Savoirs*, *50*(2), 129–153. https://doi.org/10.3917/savo.050.0129
8. Buhl, J. L. (2014). La Déclaration sociale nominative (DSN) et l'accès aux droits. *Regards*, (2), 135-140.
9. Bullich, V. (2018). La « plateformisation » de la formation. *Distances et médiations des savoirs*, *21*.
10. Bullich, V. (2022). Plateforme, plateformiser, plateformisation: le péril des mots qui occultent ce qu'ils nomment. *Questions de communication*, *40*, 47-70.
11. Caillaud, P. (2023). La formation est-elle devenue un objet de consommation?. *Éducation permanente*, (1), 137-146. https://doi.org/10.3917/edpe.234.0137
12. Hansen, J. D., et Reich, J. (2015). Democratizing education? Examining access and usage patterns in massive open online courses. *Science*, *350*(6265), 1245-1248.

13. Isaac, H. et Volle, P. (2008). *E-commerce : de la stratégie à la mise en œuvre opérationnelle*. Paris : Pearson Education France.
14. Jaumont, L. (2021). La montée en charge du nouveau Compte personnel de formation (CPF) : quels impacts sur la demande de formation ? Questions Politiques Sociales - Les brèves n°9, septembre 2021, disponible sous : https://politiques-sociales.caissedesdepots.fr/qps-les-breves-ndeg9
15. Langevin, G. (2022). Formations financées par le CPF des salariés en contrats courts ou au Smic, en 2019 et en 2020. Questions Politiques Sociales – Les études n°36, février 2022, disponible sous : https://politiques-sociales.caissedesdepots.fr/qps-les-etudes-ndeg36
16. Luttringer, J. M., et Soldini, D. (2018). L'ambition «universaliste» du nouveau droit de la formation tout au long de la vie. *Revue française des affaires sociales*, (4), 39-52. https://doi.org/10.3917/rfas.184.0039
17. Mesnard, O. (2018). La dépense nationale pour la formation professionnelle continue et l'apprentissage en 2015, Dares résultats, n°034 ; Dares, juillet 2018. https://dares.travail-emploi.gouv.fr/IMG/pdf/2018-034.pd
18. Perez, C. (2019). Avec le Compte Personnel de Formation: L'avènement d'une logique marchande et désintermédiée. *Savoirs, 50*(2), 87–100. https://doi.org/10.3917/savo.050.0087
19. Perez, C., et Vourc'h, A. (2020). *L'individualisation des dispositifs d'accès à la formation: l'exemple français du Compte Personnel de Formation* (No. 245). OECD Publishing. https://shorturl.at/szHVZ
20. Robinson, A. C., Anderson, C. L., et Quinn, S. D. (2020). Evaluating geovisualization for spatial learning analytics. *International Journal of Cartography*, *6*(3), 331-349.

Évaluation de différents critères pour la formation de groupes dans le contexte du pair programming distribué : expérimentation dans l'enseignement supérieur

José Colin[1], Sébastien Hoarau[2], and Julien Broisin[1]

[1] IRIT, Université de Toulouse, Toulouse, France
[2] Laboratoire d'Informatique et de Mathématiques, Université de la Réunion, Saint-Denis, France

Résumé. La formation des groupes est une problématique importante de l'apprentissage collaboratif, car des groupes formés de manière appropriée peuvent conduire à de meilleures interactions entre les étudiants et améliorer leur apprentissage. Cette problématique s'applique également à la formation de binômes dans le contexte d'activités d'apprentissage de pair programming. Parmi les méthodes automatiques existantes utilisées pour le pair programming, beaucoup impliquent d'avoir à disposition un certain nombre d'informations personnelles sur les étudiants, ce qui peut complexifier la mise en place de ce genre de dispositif. Sur la base de cette limite, nous proposons une méthode automatique de formation des binômes intégrée à une application de distributed pair programming, et qui s'appuie non seulement sur certaines caractéristiques personnelles des étudiants obtenues avec des données auto-rapportées, mais aussi sur les interactions antérieures des étudiants avec l'application collectées sous forme de traces. Cette méthode a été expérimentée dans le cadre d'un cours d'introduction à la programmation, dans lequel 133 sessions de travail en pair programming ont été réalisées sur notre application de distributed pair programming. Les critères de formation des binômes utilisés par le système pendant cette expérimentation étaient l'expérience en programmation, l'auto-efficacité, le genre et l'engagement. L'expérimentation montre que les binômes homogènes en auto-efficacité ont une compatibilité perçue et une répartition du travail perçue significativement meilleures que les binômes hétérogènes, et que les binômes hétérogènes en experience de programmation ont une performance académique significativement meilleure que les binômes homogènes.

Mots-clés : Apprentissage de l'informatique ; Pair programming distribué ; Formation de groupes ; Conception et évaluation

Abstract. Group formation is an important issue in collaborative learning, as appropriately formed groups can lead to better interactions between students and enhance their learning. This issue also applies to the context of pair programming learning activities. Many of the existing

automatic methods used for pair programming require a certain amount of personal information about the students, which can make setting up this kind of system very complex. Based on these limitations, we present an automatic pair formation method integrated to a distributed pair programming application, using personal characteristics of the students as formation criteria, obtained with self-reported data but also with trace data from the students. We experimented this method as part of an introductory programming course, in which 133 peer programming sessions were carried out on our distributed pair programming application. The criteria for pair formation used by the system during this experiment were programming experience, self-efficacy, gender and engagment. The experiment shows that homogeneous pairs in self-efficacy are significantly better than heterogeneous pairs, both in terms of perceived compatibility and perceived work distribution, and that the heterogeneous pairs in programming experience have a significantly higher academic performance than homogeneous pairs.

Keywords: Computer Science Education ; Distributed Pair Programming ; Group Formation ; Tool Design and Evaluation

1 Introduction

Le pair programming (PP) est une méthode de développement impliquant deux programmeurs : le *Driver*, et le *Navigator*. Le *Driver* travaille activement sur la tâche de programmation et rédige le code, tandis que le *Navigator* observe et apporte son aide au *Driver* en mettant en avant des améliorations possibles. Appliqué à l'enseignement, le PP a de multiples impacts positifs sur les étudiants, comme une plus grande confiance dans la qualité du code, de meilleures compétences en résolution de problèmes [7], ainsi qu'une plus grande satisfaction perçue des étudiants [17]. Dans un contexte où les deux programmeurs sont éloignés géographiquement, le *distributed pair programming* (DPP) permet de réaliser des activités de PP à distance. Le DPP se déroule dans un environnement virtuel où un éditeur de code est partagé entre les deux programmeurs, garantissant que seul le programmeur dans le rôle de *Driver* peut modifier le code source. L'environnement possède également des fonctionnalités pour exécuter du code et permettre au *Driver* et *Navigator* d'inverser leurs rôles. La communication peut se faire de manière textuelle ou vocale.

Pour le PP et le DPP, comme pour l'apprentissage collaboratif de manière générale, la formation des binômes est une problématique importante car elle a un impact significatif sur l'apprentissage et la participation des apprenants [9]. Une solution simple et étudiée dans différents travaux [22,10] consiste à laisser les étudiants choisir leurs binômes : ces études montrent qu'ils ont alors tendance à choisir en fonction de leurs affinités [22]. Cette solution donne un taux de compatibilité perçue important pour les binômes (plus de 95% pour [22] et [10]). Il est également possible que l'enseignant soit en charge de la formation des binômes [12]. Il se base alors sur la connaissance de ses étudiants pour former les

binômes qu'il juge les plus pertinents par rapport aux objectifs d'apprentissage qu'il a fixés.

Cependant, ces solutions ne sont pas toujours applicables. Dans l'enseignement supérieur, les étudiants de première année, souvent en grand nombre, ne se connaissent pas nécessairement. Cela rend difficile la formation des binômes en fonction de l'affinité, et le nombre important d'étudiants rend la tâche complexe et chronophage pour l'enseignant. Dans le cas des MOOCs, les mêmes problèmes se posent. Les étudiants ont encore moins d'opportunités d'apprendre à se connaître, et les enseignants ne sont pas forcément amenés à connaître leurs étudiants. Pourtant, la formation des binômes reste un enjeu majeur pour l'efficacité du PP.

Dans cet article, nous proposons une méthode automatique de formation des binômes dans le cas du DPP. Cette méthode cherche, à partir d'un nombre important d'étudiants, à effectuer des appariements en fonction de leur profil. En particulier, nous étudions leur expérience de programmation, leur sentiment d'auto-efficacité, leur genre, et leur engagement antérieur dans des activités de PP. L'objectif est d'étudier et de comparer l'impact de ces caractéristiques sur la compatibilité perçue par les étudiants avec leur partenaire, la répartition du travail perçue par les étudiants au sein du binôme, ainsi que sur la performance et l'engagement des étudiants dans l'activité de DPP. Notre approche consiste à constituer, pour une même caractéristique du profil des apprenants, des groupes homogènes d'une part et des groupes hétérogènes d'autre part.

Après un état de l'art sur la formation des groupes dans le domaine général de l'apprentissage collaboratif supporté par ordinateur, et du PP en particulier, nous présentons notre système de formation des binômes et les détails de l'algorithme qui le met en œuvre au sein d'une application existante de DPP [3]. L'évaluation de ce système dans le cadre d'une expérimentation menée dans l'enseignement supérieur et les résultats obtenus sont ensuite discutés. Finalement, nous donnons les principales conclusions de ces travaux et les pistes de recherche que nous souhaitons explorer pour améliorer notre système de formation des groupes.

2 État de l'art

Cette section présente dans un premier temps les approches de formation de groupes proposées dans le domaine du Computer Supported Collaborative Learning (CSCL), puis plus spécifiquement dans le cas du pair programming.

2.1 Formation des groupes dans le domaine du CSCL

La formation des groupes dans le domaine du CSCL fait référence aux stratégies, techniques et méthodes mises en oeuvre pour regrouper des individus en fonction de critères afin de créer des groupes qui conduisent les apprenants à de meilleures interactions et qui maximisent l'apprentissage [1]. Dans leur revue de littérature, Borges et al. [1] montrent que sur 106 études portant sur

la formation des groupes en CSCL, les principaux critères utilisés portent sur le profil des étudiants (niveau de connaissance, compétences, genre, etc.) et les préférences des étudiants et/ou des enseignants. Ils montrent également que ces critères sont utilisés pour former les groupes selon deux approches : homogène et hétérogène. L'approche homogène cherche à grouper des étudiants similaires pour un ou plusieurs critères donnés, tandis que l'approche hétérogène cherche à grouper des étudiants différents. L'approche hétérogène est plus étudiée (84 études sur 106) que l'approche homogène (36 études sur 106).

Lorsque le nombre d'étudiants ou le nombre de critères à prendre en compte est élevé, la formation des groupes peut devenir une tâche complexe à réaliser sans support informatique. Pour pallier ce problème, on peut, à la place d'une approche manuelle où l'enseignant et/ou les étudiants sont en charge de former les groupes, passer à une approche automatique. Pour cela, De nombreuses études proposent différentes méthodes algorithmiques [4]. Les algorithmes utilisant des modèles probabilistes sont les plus utilisés. Ils permettent notamment de gérer un nombre important de variables.

2.2 Études existantes sur la formation des binômes en pair programming

Dans le cas du pair programming, comme dans celui du CSCL, les études existantes sur la formation des binômes utilisent soit des méthodes manuelles en laissant les étudiants ou les enseignants choisir les binômes [22,12], soit des méthodes automatiques [2,5]. Les critères de formation des binômes portant sur le niveau de connaissance et de compétence des étudiants sont largement utilisés [20,5,6,21], ainsi que les approches homogène et hétérogène. Par ailleurs, l'auto-efficacité est également utilisée comme critère de formation des binômes dans l'étude proposée par Tsai et al. [19].

Plusieurs études [8,12,21,2] utilisent les traits de personnalité des étudiants comme critère pour la formation des binômes en s'appuyant sur le modèle Myers-Briggs Type Indicator (MBTI) [14]. Choi et al. [2] ont montré que des binômes diversifiés en termes de personnalité font preuve d'une plus grande productivité que les binômes similaires ou opposés. Cependant, la mesure des traits de personnalité nécessite de faire passer de longs questionnaires tels que le questionnaire MBTI, ce qui n'est pas applicable à tous les contextes.

D'autre part, pour évaluer l'efficacité d'une méthode de formation de binômes, plusieurs indicateurs sont utilisés dans la littérature. La compatibilité perçue entre les deux membres du binôme est mesurée via des questionnaires post-expérimentation dans plusieurs études [12,6,5,21]. La performance des binômes dans les tâches de PP est évaluée en mesurant le temps nécessaire au binôme pour réaliser la tâche de programmation [8] et/ou en évaluant directement le code produit par le binôme [5,8].

Parmi les recherches existantes, nous n'avons pas trouvé d'études exploitant les traces d'interaction ou les learning analytics pour la formation des binômes. Par exemple, l'utilisation de l'engagement pour la formation des binômes, mesuré à partir des interactions de l'utilisateur, n'a pas, à notre connaissance, été étudiée

dans la littérature. Exploiter les traces d'interaction peut permettre d'obtenir des mesures plus objectives sur le comportement réel des étudiants en PP qu'avec des données auto-rapportées.

Dans la suite de cet article, nous essayons donc d'apporter des éléments de réponse à la question de recherche (QR) suivante : Comment la formation automatique de groupes homogènes et hétérogènes en termes d'auto-efficacité des apprenants, d'expérience en programmation, d'engagement antérieur dans des activités de pair programming, et de genre, affecte la compatibilité et la répartition du travail perçues, l'engagement, et la performance des apprenants dans le DPP ?

3 Système de formation des binômes

Cette section introduit d'abord notre application de distributed pair programming existante, puis expose le nouveau système de formation des binômes intégré dans cette application.

3.1 L'application de distributed pair programming

Le système de formation automatique des binômes que nous proposons est intégré directement dans notre application de distributed pair programming [3]. C'est une application web permettant de réaliser des travaux pratiques en Python. Au niveau de l'interface visible par les étudiants, toutes les fonctionnalités importantes pour la mise en œuvre du DPP sont implémentées : un éditeur de code partagé, un système de rôle pour différencier le *Driver* et le *Navigator* et permettre les changements de rôle, une console Python pour l'exécution du code, et une messagerie textuelle instantanée. Pour l'enseignant, l'application met à disposition une interface pour créer et modifier des activités pédagogiques. Un certain nombre de traces d'interactions sont collectées : connexion et déconnexion des utilisateurs ; demandes de changement de rôle et les réponses associées ; captures du code source Python produit ; interactions avec la console Python ; messages échangés dans le canal textuel. Elles sont horodatées et formatées en format JSON.

L'application a déjà été utilisée dans le cadre d'expérimentations visant à évaluer son utilisabilité [3]. Elle est accessible avec le protocole LTI [13], ce qui permet aux étudiants d'y accéder depuis une plate-forme d'apprentissage telle que Moodle.

3.2 Fonctionnement général du système de formation des binômes

Pour démarrer une session de travail en PP en utilisant le système de formation des binômes, l'enseignant commence par créer une activité sur l'application, en renseignant une consigne pour un exercice de programmation. Les étudiants se connectent ensuite simultanément sur cette activité. Ils sont alors placés dans

une file d'attente, qui se vide à mesure que les binômes sont formés et redirigés vers l'environnement de DPP.

Le système dispose des informations nécessaires sur le profil des étudiants pour effectuer les calculs de score d'appariement. À intervalle régulier, le système va prendre l'ensemble des combinaisons d'étudiants possibles dans la file d'attente et calculer pour chacune un score d'appariement selon une méthode de calcul du score d'appariement choisie. La combinaison ayant le plus grand score est alors sélectionnée pour former un binôme. Les 2 étudiants concernés quittent la file d'attente, puis sont redirigés vers l'activité de DPP. Cette opération est répétée jusqu'à ce que la file d'attente soit vide (ou jusqu'à ce qu'elle ne contienne plus qu'un seul étudiant). L'implémentation de ce système est décrite dans l'algorithme 1.

Algorithme 1 Algorithme de formation des binômes

$etudiants \leftarrow fileAttente$ ▷ L'ensemble des étudiants dans la file d'attente
while $size(etudiants) \geq 2$ **do**
  $maxScore \leftarrow 0$
  $meilleureCombinaison \leftarrow \emptyset$
  for $(etudiant1, etudiant2) \in etudiants \times etudiants$ **do**
    if $etudiant1 \neq etudiant2$ **then**
      $scoreAppariement \leftarrow calculScoreAppariement(etudiant1, etudiant2)$
      if $scoreAppariement > maxScore$ **then**
        $maxScore \leftarrow scoreAppariement$
        $meilleureCombinaison \leftarrow \{etudiant1, etudiant2\}$
  $formerBinome(meilleureCombinaison)$
  $etudiants \leftarrow etudiants \setminus meilleureCombinaison$

Il est également possible d'introduire dans ce système un seuil pour le score d'appariement. Dans ce cas-là, si parmi toutes les combinaisons d'étudiants, aucune ne dépasse ce seuil, alors aucun binôme n'est formé, et le système attend pendant une durée prédéfinie avant de recommencer à chercher des binômes. Cette attente permet d'éviter de former des binômes trop rapidement et de laisser l'opportunité à la file d'attente de se remplir afin de trouver de meilleures combinaisons d'étudiants. Si, après plusieurs attentes consécutives, aucune combinaison ne dépasse toujours pas le seuil, il est alors progressivement diminué jusqu'à ce que la situation se débloque. Cet ajout est pertinent dans le cas où les étudiants ne se connectent pas forcément tous en même temps à l'application. Pour l'expérimentation présentée dans cette étude, nous n'avons pas utilisé de système de seuil.

3.3 Calcul du score d'appariement

Le calcul du score d'appariement se fait selon une méthode choisie. Une méthode est associée à un critère de formation et elle a un objectif : maximiser l'homogénéité entre les deux étudiants sur ce critère ou maximiser l'hétérogénéité sur ce

critère. Pour une méthode qui maximise l'homogénéité, le score d'appariement s'obtient avec la formule suivante :

$$scoreAppariement = \frac{1}{\max(0.01, \mid critereEtudiant1 - critereEtudiant2 \mid)}$$

Pour une méthode qui maximise l'hétérogénéité, le score d'appariement s'obtient avec la formule suivante :

$$scoreAppariement = \mid critereEtudiant1 - critereEtudiant2 \mid$$

4 Méthode de conduite de la recherche

4.1 Configuration expérimentale

L'étude que nous avons menée a été réalisée à travers notre application de DPP [3] décrite dans la section 3.1, dans le cadre d'un cours d'introduction au langage Python à l'Université de la Réunion. Cette expérimentation a impliqué une cohorte d'étudiants inscrits en première année de licence informatique, et une cohorte d'étudiants inscrits en première année de licence de mathématiques. Chaque cohorte a utilisé notre plate-forme pendant deux séances de cours magistraux dispensées en présentiel par le même enseignant. Le tableau 1 donne certaines données démographiques sur les étudiants ayant participé à cette expérimentation.

Table 1. Données démographiques sur les participants

Nombre total	Nombre de L1 Informatique	Nombre de L1 Mathématiques	Nombre de filles	Nombre de garçons	Moyenne d'âge
102	82	20	8	94	19,6

Durant la première séance de cours magistral, l'enseignant a tout d'abord présenté notre plate-forme afin d'exposer ses principales fonctionnalités. Ensuite, les étudiants ont disposé de 20 minutes pour réaliser une activité de PP conçue par l'enseignant. Durant la deuxième séance de cours magistral, les étudiants ont participé à deux activités de PP, chacune d'une durée de 40 minutes.

Chaque activité de PP était centrée sur des notions vues précédemment dans le cadre du cours d'introduction au langage Python. Voici un exemple de consigne pour une activité de PP en Python : "Ecrire une fonction ‘echange‘, qui intervertit deux cases d'un tableau donné. ‘nombres = [45, 34, 12, 23, 57]‘ echange(nombres,2,4)‘ doit modifier nombres en : ‘[45, 34, 57, 23, 12]‘" Elles étaient accessibles depuis la plate-forme d'apprentissage Moodle de l'Université Anonyme. Ainsi, les étudiants étaient automatiquement connectés sur notre plate-forme via le protocole LTI (voir section 3.1). Lors de leur première connexion

à la plate-forme (i.e., lors de la réalisation de la première activité), les étudiants devaient remplir le questionnaire pré-expérimental Q1 concernant leur expérience en programmation et leur sentiment d'auto-efficacité. Après avoir réalisé chacune des trois activités, ils devaient remplir le questionnaire post-expérimentation Q2 concernant la répartition du travail perçue au sein de leur binôme ainsi que la compatibilité perçue avec leur partenaire.

Au début de chaque activité, les binômes ont été formés par le système de formation automatique décrit dans la section 3.2. Le système s'est appuyé sur différents critères de formation (i.e., l'expérience en programmation des apprenants ; leur sentiment d'auto-efficacité ; leur genre ; leur engagement antérieur dans des activités de PP), en essayant d'équilibrer autant que possible le nombre de binômes formés selon chaque critère de formation, en fonction des données disponibles sur les étudiants. Les étudiants se sont donc vus attribuer un binôme localisé à n'importe quel endroit de l'amphithéâtre, et devaient communiquer via le canal textuel intégré à l'application. Pendant la première séance, les critères de formation utilisés étaient l'expérience en programmation, l'auto-efficacité et le genre. Pour la deuxième séance, les critères utilisés étaient l'engagement antérieur, l'auto-efficacité et le genre.

4.2 Données collectées

Les données collectées sont anonymisées. Seul l'identifiant Moodle transmis par LTI est collecté, ne permettant pas de révéler l'identité des étudiants.

Données auto-rapportées. Le questionnaire pré-expérimental Q1, dispensé via notre plate-forme DPP lors de la première connexion seulement et disponible sur recherche.data.gouv.fr., a pour objectif de mesurer trois critères utilisés dans le système de formation des binômes : l'expérience en programmation des apprenants, leur auto-efficacité et leur genre. L'expérience en programmation est obtenue à partir du questionnaire proposé par Siegmund et al. [18], traduit en français. Ce questionnaire comprend 15 questions au total, parmi lesquelles 8 questions à choix uniques et 7 questions de Likert à 5 niveaux. Le **score d'expérience de programmation** est obtenu en calculant la moyenne des réponses à ces questions. Nous définissons, pour chaque question à choix unique, une méthode pour obtenir un score. Le détail du calcul est disponible sur recherche.data.gouv.fr. Nous avons utilisé le questionnaire proposé par Tsai et al. [19], traduit en français, pour mesurer l'auto-efficacité des apprenants. Ce questionnaire est constitué de 16 questions de Likert à 6 niveaux. Le **score d'auto-efficacité** est obtenu en calculant la moyenne des réponses formulées par le répondant. Enfin, **le genre** est obtenu à travers 1 question à choix unique.

Le questionnaire post-expérimentation Q2, également dispensé via notre plate-forme DPP mais cette fois à la fin de chaque activité de PP, vise à mesurer leur perception de la répartition du travail avec leur partenaire, ainsi que la compatibilité perçue par les étudiants au sein de leur binôme . Ce questionnaire apparaît en Annexe A.. Nous nous sommes appuyés sur le questionnaire de Williams et

al. [21], traduit en français, qui vise à mesurer lors de situations de PP, (i) la répartition du travail perçue par les étudiants à l'aide de 3 questions de Likert à 5 niveaux, et (ii) la compatibilité perçue à l'aide d'1 question de Likert à 3 niveaux. Le **score de répartition du travail perçue *au sein du binôme*** est calculé en faisant la moyenne des réponses fournies par les 2 membres du binôme aux 3 questions R1, R2 et R3. Le **score de compatibilité perçue *au sein du binôme*** est calculé en suivant la même démarche, en calculant la moyenne de la réponse des 2 membres à la question C1.

Données quantitatives. Comme mentionné dans la section 3.1, notre application de DPP collecte différentes traces d'interactions. Ces données sont utilisées pour calculer 2 indicateurs : la performance académique du binôme et son engagement dans l'activité de PP. La **performance académique** correspond à un score obtenu à partir du code source final produit par le binôme par rapport à l'activité proposée par l'enseignant. Elle correspond à l'évaluation manuelle du code source selon un barème établi par l'enseignant ayant proposé l'activité. Afin d'éviter les biais de notation, un seul et unique correcteur a procédé à l'évaluation de l'ensemble des codes sources. Ainsi, pour chacune des 3 activités, chaque binôme a obtenu un score de performance académique.

L'engagement du binôme dans l'activité de PP est obtenu en combinant différentes variables calculées pour chaque membre du binôme :

- Le nombre de caractères insérés/supprimés (*nbChar*). Cette valeur est obtenue en comparant les différences successives entre chaque capture du contenu de l'éditeur de code. Pour chaque modification du code, la différence en nombre de caractères avant et après la modification est ajoutée au total de l'étudiant auteur de la modification. Cette différence est obtenue avec l'algorithme de Myers [15]. Cette façon de faire permet de bien distinguer la contribution des deux membres du binôme.
- Le nombre d'exécutions du code et de commandes tapées dans la console Python (*nbCommand*).
- Le nombre de demandes de changement de rôle, et de réponses à une demande de changement de rôle (*nbRoleChange*).
- Le nombre de messages écrits dans le canal de communication textuel (*nbMessage*).

Ces variables sont ensuite normalisées en divisant par les valeurs maximales obtenues dans le jeu de données lors de la première séance de l'expérimentation ($nbCharMax = 920$, $nbCommandMax = 19$, $nbRoleChangeMax = 5$, $nbMessageMax = 31$). Ainsi, pour chaque étudiant, nous calculons un score d'engagement compris entre 0 et 1 en appliquant la formule suivante :

$$engagement = \frac{(\frac{nbChar}{920} + \frac{nbCommand}{19} + \frac{nbRoleChange}{5} + \frac{nbMessage}{31})}{4}$$

Finalement, le **score d'engagement du binôme** dans l'activité de PP est obtenu en calculant la moyenne de l'engagement individuel des deux membres du

binôme. Notons ici que le score d'engagement individuel de la formule ci-dessus a également été utilisé, lors de la deuxième séance, pour le critère de formation d'engagement antérieur.

4.3 Analyse des données

Pour répondre à notre question de recherche concernant l'impact de binômes homogènes et hétérogènes en termes d'auto-efficacité, d'expérience en programmation, de genre et d'engagement antérieur dans des activités de PP sur la compatibilité et la répartition du travail perçues, l'engagement et la performance des apprenants, nous avons rapporté les moyennes, déviations standard et médianes. Nous avons ensuite cherché à identifier si des différences significatives existent entre le groupe homogène et le groupe hétérogène sur chacun des critères de formation utilisés. Nous avons utilisé des tests U de Mann-Whitney car nos données n'étaient pas normalement distribuées. Ces tests ont été réalisés pour les 4 variables *perception de la répartition du travail*, *perception de la compatibilité*, *performance académique*, et *engagement dans l'activité de pair programming*.

Pour classer les binômes dans les groupes homogène et hétérogène dans notre analyse, nous avons placé, pour chaque critère de formation, la moitié des binômes ayant la différence la plus faible sur le critère dans le groupe hétérogène, et l'autre moitié dans le groupe homogène. Pour le critère du genre, seuls 8 binômes hétérogènes (homme/femme) ont pu être formés, les deux groupes ne sont donc pas de même taille.

5 Résultats

Au total, 133 sessions de travail de DPP, avec des binômes différents à chaque fois, ont été prises en compte dans notre étude. 51 et 82 sessions ont respectivement été réalisées pendant la première et la deuxième séance de cours magistral. Le tableau 2 donne le nombre de binômes créés lors de l'expérimentation pour chaque critère de formation.

Table 2. Nombre de binômes par critère de formation

	Expérience en programmation	Auto-efficacité	Genre	Engagement antérieur
Nombre de binômes	21	50	35	27

Le nombre de binômes formés n'est pas le même pour chacun des critères. En effet, le critère "expérience en programmation" n'a été utilisé que pendant la première séance de cours magistral, où un nombre moins important de sessions de travail ont été réalisées. Aussi, le critère "engagement antérieur" n'a été utilisé que lors de la deuxième séance de cours magistral. En effet, ce critère ne peut être

calculé automatiquement que lorsqu'un étudiant a réalisé au moins une activité de PP sur notre plate-forme (voir section 4.3).

Les figures 1, 2, 3 et 4 présentent les moyennes obtenues pour les indicateurs de compatibilité perçue, de performance académique et d'engagement en fonction des différentes méthodes de formation des groupes utilisées. La figure 1 ne fait pas apparaître de différences importantes dans les moyennes de compatibilité perçue entre les groupes homogènes et hétérogènes en expérience de programmation et en genre, mais fait apparaître des différences plus importantes pour l'auto-efficacité favorisant le groupe homogène et pour l'engagement favorisant le groupe hétérogène. La figure 2 montre de légères différences pour le genre et l'engagement, et une différence un peu plus importante pour l'auto-efficacité dans la répartition du travail perçue moyenne. Comme pour la compatibilité perçue, c'est le groupe homogène qui obtient de meilleurs résultats pour l'auto-efficacité. La figure 3 montre une différence significative pour l'expérience de programmation en faveur du groupe hétérogène dans la performance académique du binôme, mais pas de différence pour l'auto-efficacité, le genre et l'engagement. Enfin, la figure 4 met en évidence des différences importantes pour l'expérience en programmation et l'engagement, et pas de différences pour l'auto-efficacité et le genre dans l'engagement du binôme.

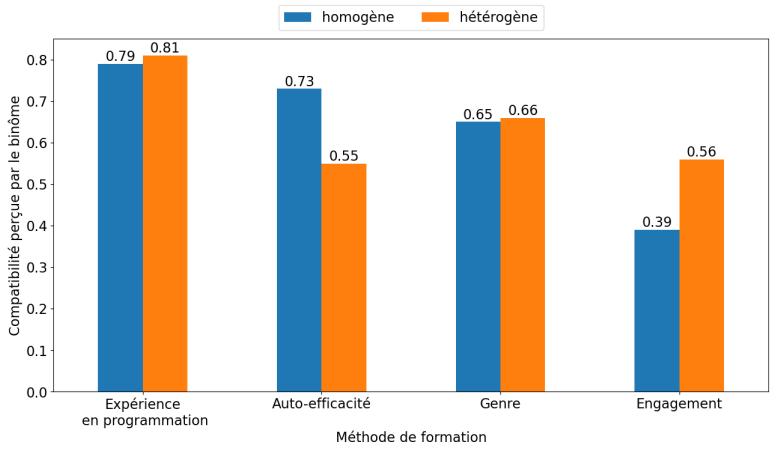


Figure 1. Compatibilité perçue moyenne des binômes en fonction des méthodes de formations utilisées

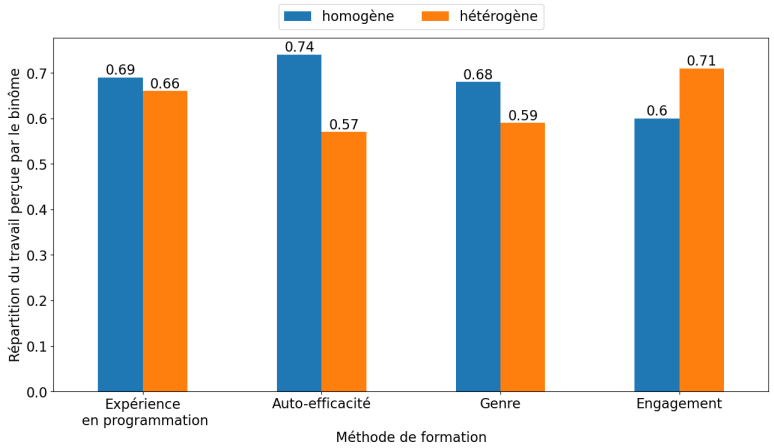


Figure 2. Répartition du travail perçue moyenne des binômes en fonction des méthodes de formations utilisées

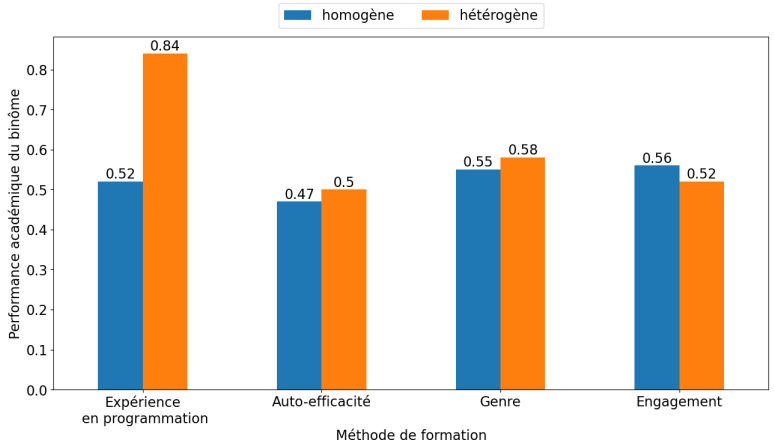


Figure 3. Performance académique moyenne des binômes en fonction des méthodes de formations utilisées

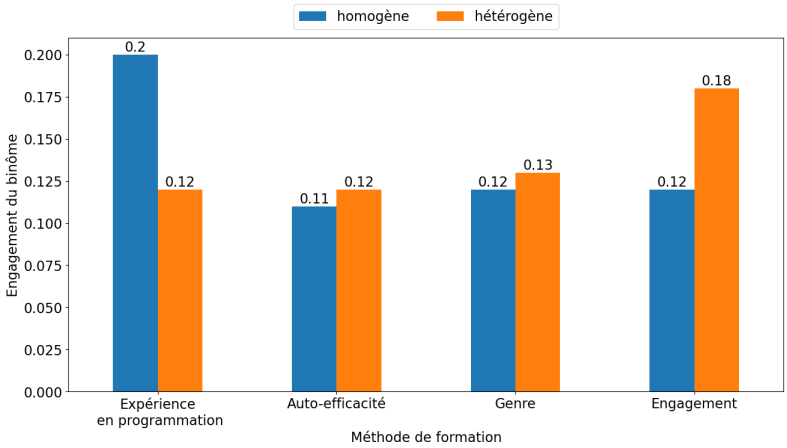


Figure 4. Engagement moyen des binômes en fonction des méthodes de formations utilisées

La table 3 présente les résultats des tests U de Mann-Whitney comparant les groupes homogènes et les groupes hétérogènes pour chaque critère de formation, ainsi que le détail des moyennes, déviations standard et médianes pour les indicateurs de compatibilité et de répartition du travail perçus, d'engagement et de performance des apprenants. Cette table fait apparaître 3 résultats montrant une différence significative ($p < .05$) entre les groupes homogènes et hétérogènes : les binômes ayant une auto-efficacité homogène ont une compatibilité perçue et une répartition du travail perçue significativement plus importante que les binômes hétérogènes, et les binômes ayant une expérience en programmation hétérogène ont une performance académique significativement meilleure que les binômes homogènes.

6 Discussion

Nous observons que les binômes avec une auto-efficacité homogène ont une compatibilité perçue significativement plus importante que les binômes avec une auto-efficacité hétérogène ($p = 0.01$). On obtient un résultat similaire pour la répartition du travail perçue ($p = 0.01$). Ce résultat va dans le sens d'une étude menée par Katira et al., dans laquelle les auteurs ont montré que les binômes ayant une compétence technique perçue homogène étaient plus compatibles que les binômes hétérogènes [12]. Ces résultats suggèrent que ce critère de formation peut être utilisé pour maximiser la répartition du travail au sein des binômes.

On peut également faire le rapprochement entre ce résultat et des études montrant une meilleure compatibilité perçue pour les binômes homogènes en compétence de programmation [5,11,20]. Cependant, notre étude ne montre pas

TABLE 3. Résultats des tests U de Mann-Whitney comparant les groupes homogènes et les groupes hétérogènes pour chaque critère de formation

	Compatibilité perçue par le binôme					Répartition du travail perçue par le binôme					Performance académique du binôme					Engagement du binôme				
Méthode de formation	M	SD	Mdn	U	p	M	SD	Mdn	U	p	M	SD	Mdn	U	p	M	SD	Mdn	U	p
Expérience en programmation	0.79	0.25	1	210	0.77	0.69	0.66	1	248.5	0.47	0.52	0.41	0.50	127.5	**0.01***	0.203	0.205	0.106	250	0.46
homogène																				
Expérience en programmation	0.81	0.25	1			0.66	0.36	0.83			0.84	0.30	1			0.120	0.085	0.132		
hétérogène																				
Auto-efficacité	0.73	0.31	1	1604	**0.01***	0.74	0.32	0.83	1634	**0.01***	0.47	0.42	0.33	1210	0.78	0.111	0.117	0.082	1148.5	0.48
homogène																				
Auto-efficacité	0.55	0.34	0.5			0.57	0.34	0.58			0.50	0.42	0.5			0.123	0.117	0.113		
hétérogène																				
Genre homogène	0.65	0.29	0.5	424.5	0.91	0.68	0.34	0.75	490	0.41	0.55	0.42	0.67	436	0.96	0.118	0.087	0.104	387	0.53
Genre hétérogène	0.66	0.30	0.5			0.59	0.37	0.71			0.58	0.31	0.58			0.125	0.087	0.115		
Engagement homogène	0.39	0.29	0.5	282	0.13	0.60	0.39	0.75	319.5	0.43	0.56	0.36	0.5	389.5	0.66	0.123	0.110	0.110	280	0.15
Engagement hétérogène	0.56	0.45	0.5			0.71	0.31	0.67			0.52	0.34	0.5			0.182	0.145	0.179		

*$p < .05$

de différence significative sur la compatibilité perçue pour le critère d'expérience en programmation. Une limite conduisant possiblement à ce résultat est discutée dans la section 7.

Pour l'engagement, malgré des écarts de moyenne importants entre le groupe homogène et hétérogène sur deux des critères de formation, les résultats des tests T indépendants ne montrent pas de différences significatives. Cela peut s'expliquer par la déviation standard importante que l'on observe sur cet indicateur.

En ce qui concerne la performance académique, les binômes hétérogènes en expérience de programmation ont obtenu des résultats significativement supérieurs aux binômes homogènes ($p = 0{,}01$). Ce résultat est cohérent avec des études suggérant qu'associer des individus ayant des expériences diverses peut conduire à de bons résultats. [16].

7 Limitations

Dans cette étude, nous avons cherché à reproduire les conditions du DPP, en éloignant physiquement les étudiants d'un même binôme. Même si les étudiants n'étaient pas côte à côte, le fait qu'ils soient présents dans la même salle de cours a pu en amener certains à vouloir communiquer directement avec leurs binômes, ce qui a pu perturber le déroulement des sessions de travail.

L'expérience en programmation a été mesurée à l'aide de réponses à un questionnaire portant sur le profil des étudiants. Or, les étudiants ayant pour beaucoup le même âge et une expérience en programmation similaire venant essentiellement de leur expérience au lycée, la déviation standard sur cet indicateur est très faible. Les groupes considérés comme homogènes ou hétérogènes sur ce critère dans notre étude ne le sont en réalité que très peu. Il aurait été préférable d'utiliser une cohorte d'étudiants plus diverse.

Le calcul de l'indicateur d'engagement a été réalisé de façon arbitraire, notamment au niveau du poids accordé à chaque variable. Il pourrait être intéressant de proposer une autre formule pour le calcul de l'indicateur d'engagement en mettant en œuvre une démarche scientifique pour déterminer les poids adéquats pour chaque variable.

Pour les binômes hétérogènes en genre, la petite taille de l'échantillon (8 binômes) rend nos résultats difficilement généralisables.

8 Conclusion

Dans cet article, nous proposons un système de formation des binômes, explorant plusieurs méthodes de formation cherchant à maximiser l'homogénéité ou l'hétérogénéité sur des caractéristiques personnelles des étudiants. Nous avons utilisé ce système dans une expérimentation où 133 sessions de travail en distributed pair programming ont eu lieu. Parmi les méthodes de formation utilisées dans cette étude, aucune ne semble être meilleure sur la totalité de nos indicateurs (compatibilité perçue, répartition du travail perçue, performance académique et

engagement). Cependant, les binômes homogènes en auto-efficacité ont obtenu des résultats significativement meilleurs sur deux indicateurs : la compatibilité et la répartition du travail perçue. Toutes les données de cette expérimentation sont disponibles en ligne sur recherche.data.gouv.fr.

Cette étude n'utilise que des méthodes utilisant un unique critère de formation. Il pourrait être intéressant, en réutilisant les mêmes critères pour la formation des binômes, d'utiliser des approches hybrides [1], utilisant plusieurs critères simultanément.

Références

1. Borges, S., Mizoguchi, R., Bittencourt, I.I., Isotani, S. : Group formation in cscl : A review of the state of the art. In : Cristea, A.I., Bittencourt, I.I., Lima, F. (eds.) Higher Education for All. From Challenges to Novel Technology-Enhanced Solutions. pp. 71–88. Springer International Publishing, Cham (2018)
2. Choi, K.S., Deek, F.P., Im, I. : Exploring the underlying aspects of pair programming : The impact of personality. Information and Software Technology **50**(11), 1114–1126 (2008). https://doi.org/https://doi.org/10.1016/j.infsof.2007.11.002, https://www.sciencedirect.com/science/article/pii/S0950584907001292
3. Colin, J., Hoarau, S., Declercq, C., Broisin, J. : Design and evaluation of a web-based distributed pair programming tool for novice programmers. In : Proceedings of the 2024 on Innovation and Technology in Computer Science Education V. 1. p. 527–533. ITiCSE 2024, Association for Computing Machinery, New York, NY, USA (2024). https://doi.org/10.1145/3649217.3653571, https://doi.org/10.1145/3649217.3653571
4. Cruz, W.M., Isotani, S. : Group formation algorithms in collaborative learning contexts : A systematic mapping of the literature. In : Baloian, N., Burstein, F., Ogata, H., Santoro, F., Zurita, G. (eds.) Collaboration and Technology. pp. 199–214. Springer International Publishing, Cham (2014)
5. Ömer Demir, Seferoglu, S.S. : The effect of determining pair programming groups according to various individual difference variables on group compatibility, flow, and coding performance. Journal of Educational Computing Research **59**(1), 41–70 (2021). https://doi.org/10.1177/0735633120949787, https://doi.org/10.1177/0735633120949787
6. Dou, W., He, W. : Compatibility and requirements analysis of distributed pair programming. In : 2010 Second International Workshop on Education Technology and Computer Science. vol. 1, pp. 467–470 (2010). https://doi.org/10.1109/ETCS.2010.367
7. Faja, S. : Pair programming as a team based learning activity : A review of research. Issues in Information Systems **XII**, 207–216 (01 2011)
8. Hannay, J.E., Arisholm, E., Engvik, H., Sjoberg, D.I. : Effects of personality on pair programming. IEEE Transactions on Software Engineering **36**(1), 61–80 (2010). https://doi.org/10.1109/TSE.2009.41
9. Isotani, S., Inaba, A., Ikeda, M., Mizoguchi, R. : An ontology engineering approach to the realization of theory-driven group formation. International Journal of Computer-Supported Collaborative Learning **4**, 445–478 (2009)

10. Jacobson, N., Schaefer, S.K. : Pair programming in cs1 : overcoming objections to its adoption. SIGCSE Bull. **40**(2), 93–96 (Jun 2008). https://doi.org/10.1145/1383602.1383643, https://doi.org/10.1145/1383602.1383643
11. Jensen, R. : A pair programming experience. The Journal of Defense Software Engineering **16**(3), 22–24 (2003), http://www.stsc.hill.af.mil/crosstalk/2003/03/jensen.html
12. Katira, N., Williams, L., Wiebe, E., Miller, C., Balik, S., Gehringer, E. : On understanding compatibility of student pair programmers. SIGCSE Bull. **36**(1), 7–11 (Mar 2004). https://doi.org/10.1145/1028174.971307, https://doi.org/10.1145/1028174.971307
13. Lti. https://www.1edtech.org/standards/lti (2024)
14. Meyers, B.I., Meyers, P.B. : Gifts differing understanding personality type (1995)
15. Myers, E.W. : Ano(nd) difference algorithm and its variations. Algorithmica **1**(1), 251–266 (Nov 1986). https://doi.org/10.1007/BF01840446, https://doi.org/10.1007/BF01840446
16. Salinas, E.Y., Williams, A.E., King, C.E. : Effect of controlling group heterogeneity on student performance in a graphical programming course. In : 2019 IEEE Frontiers in Education Conference (FIE). pp. 1–8 (2019). https://doi.org/10.1109/FIE43999.2019.9028593
17. Salleh, N., Mendes, E., Grundy, J. : Empirical studies of pair programming for cs/se teaching in higher education : A systematic literature review. IEEE Transactions on Software Engineering **37**(4), 509–525 (2011)
18. Siegmund, J., Kästner, C., Liebig, J., Apel, S., Hanenberg, S. : Measuring and modeling programming experience. Empirical Software Engineering **19**(5), 1299–1334 (Oct 2014). https://doi.org/10.1007/s10664-013-9286-4, https://doi.org/10.1007/s10664-013-9286-4
19. Tsai, M.J., Wang, C.Y., Hsu, P.F. : Developing the computer programming self-efficacy scale for computer literacy education. Journal of Educational Computing Research **56**(8), 1345–1360 (2019). https://doi.org/10.1177/0735633117746747, https://doi.org/10.1177/0735633117746747
20. Van Toll III, T., Lee, R., Ahlswede, T. : Evaluating the usefulness of pair programming in a classroom setting. In : 6th IEEE/ACIS International Conference on Computer and Information Science (ICIS 2007). pp. 302–308 (2007). https://doi.org/10.1109/ICIS.2007.96
21. Williams, L., Layman, L., Osborne, J., Katira, N. : Examining the compatibility of student pair programmers. In : AGILE 2006 (AGILE'06). pp. 10 pp.–420 (2006). https://doi.org/10.1109/AGILE.2006.25
22. Xinogalos, S., Satratzemi, M., Chatzigeorgiou, A., Tsompanoudi, D. : Student perceptions on the benefits and shortcomings of distributed pair programming assignments. In : 2017 IEEE Global Engineering Education Conference (EDUCON). pp. 1513–1521 (2017). https://doi.org/10.1109/EDUCON.2017.7943050

Annexe A. Questionnaire Q2

- R1. Votre partenaire de pair programming a-t-il fait sa juste part du travail ?
- R2. Votre partenaire a-t-il suivi le modèle de pair programming (en alternant les rôles de driver et de navigator) ?
- R3. Votre partenaire de pair programming a-t-il contribué à la réalisation de l'exercice ? A-t-il coopéré ?
- C1. D'après vous, à quel point vous et votre partenaire étaient compatibles ?

Rendre explicable un moteur de recommandations pédagogiques pour permettre une co-évolution système-enseignant.e

Chloé Conrad[1], Rémi Venant[2], Nathalie Guin[1], and Marie Lefevre[1]

[1] Universite Claude Bernard Lyon 1, CNRS, INSA Lyon, LIRIS, UMR5205, 69622 Villeurbanne, France
chloe.conrad@univ-lyon1.fr,
[2] Le Mans Université, LIUM, 53000 Laval, France

Résumé Dans le cadre du projet ComPer, l'approche par compétences a permis de proposer aux équipes pédagogiques et aux apprenant.e.s un système de recommandations de ressources pédagogiques. La contribution présentée dans cet article consiste en l'apport de deux types d'explications aux enseignant.e.s sur le fonctionnement de ce système de recommandation, afin de leur faciliter son paramétrage. Ces explications sont générées à partir des traces d'exécution du système. L'utilisabilité de l'interface de présentation des explications et la pertinence de ces dernières pour appréhender le système ont été mises à l'essai auprès de membres des équipes pédagogiques utilisatrices des outils ComPer.

Keywords: Explicabilité · Intelligence artificielle pour l'éducation · Approche par compétences · Système de recommandation.

Abstract. The aim of the ComPer project is to provide digital tools that enable teachers to implement a Competency-Based Approach and personalize learning. This is achieved through a configurable system for recommending teaching resources. The contribution presented in this paper involves providing teachers with two types of explanations regarding how this recommendation system works, in order to facilitate its configuration. These explanations are generated from the system's execution traces. The usability of the interface used to present the explanations, as well as their relevance for understanding the system, have been evaluated with members of the teaching teams who use the ComPer tools.

Keywords: Explainability · Artificial Intelligence for Education · Competency-Based Approach · Recommendation System.

1 Introduction

L'Approche par Compétences (APC), associée à des formes de pédagogies actives, fait l'objet d'un soutien institutionnel affirmé de la part des autorités politiques. L'APC met l'accent sur les objectifs et les résultats observables de l'activité d'apprentissage afin de permettre aux apprenant.e.s de développer des compétences cohérentes avec les besoins du marché du travail [6]. Cependant, cette approche reste peu mise en place, car les enseignant.e.s ne disposent pas de ressources pour la mettre en œuvre.

Dans ce contexte, le projet ANR ComPer visait à développer des outils numériques permettant la mise en œuvre de l'APC. Il a permis de proposer un cadre permettant aux enseignant.e.s de concevoir des référentiels de compétences et d'y associer des activités pédagogiques qui permettent de travailler les compétences définies et qui conviennent à leur pratique d'enseignant.e.s (COM, compétences) puis de recommander automatiquement parmi ces activités, celles adaptées aux compétences de chacun de leurs apprenant.e.s (PER, personnalisation). Différents modèles et outils ont été réalisés : un méta-modèle de référentiels de compétences, un moteur d'IA permettant d'analyser les traces d'activités des apprenant.e.s afin de diagnostiquer leur maîtrise des compétences travaillées et ainsi constituer un profil de compétences, et un moteur d'IA permettant de recommander des ressources pédagogiques, moteur paramétrable par les enseignant.e.s. Ce système de recommandations exploite, dans chaque enseignement, un référentiel de compétences conforme au méta-modèle permettant la mise en place de l'APC en proposant une structuration d'un domaine en compétences à acquérir [15].

Le projet ANR Tea-TIME s'inscrit dans la continuité du projet ComPer avec pour objectif d'étendre les outils de ce dernier pour favoriser leur appropriation par les utilisateur.ice.s finaux : enseignant.e.s et apprenant.e.s. Le travail décrit dans cet article se situe dans ce cadre et s'intéresse en particulier à l'appropriation du système de recommandations par les enseignant.e.s *via* son paramétrage.

Notre article s'articule comme suit. Dans la section 2, nous explorerons l'état de l'art des méthodes d'explicabilité en IA, notamment en éducation. Dans la section 3, nous présenterons les outils du projet ComPer. Notre contribution, un moteur d'explications pour le système de recommandations et la restitution de ces explications sur l'outil de paramétrage de ce système, sera décrite en section 4. Une mise à l'essai auprès d'équipes pédagogiques sera présentée en section 5.

2 État de l'art

2.1 Intelligence artificielle et explicabilité

L'un des enjeux centraux de l'utilisation de systèmes de recommandations en éducation réside dans la nécessité d'un haut degré de confiance et de compréhension du système de la part des utilisateurs : apprenant.e.s et enseignant.e.s. Une solution à cette problématique est l'apport d'explications sur le fonctionnement

du système de recommandations [18]. Ces systèmes étant des intelligences artificielles, nous nous intéressons à la littérature autour de l'apport d'explications en IA afin d'identifier les problématiques et les approches existantes.

D'un côté, la littérature identifie deux grandes approches pour rendre les systèmes d'IA explicables : l'apport d'explications locales et l'apport d'explications globales. La première consiste à expliquer des décisions ou des prédictions individuellement, alors que la deuxième a pour objectif d'expliquer le modèle et sa logique comme un tout [17].

D'un autre côté, les systèmes d'IA sont souvent classés en deux catégories : **les boîtes blanches et les boîtes noires**. Cette distinction repose sur la transparence de leur fonctionnement. En effet, un système boîte noire est un système dont le fonctionnement est difficilement accessible par la complexité de son architecture et le grand nombre de ses paramètres [7]. À l'inverse, dans un système dit boîte blanche, les informations sur son fonctionnement sont accessibles. Cette typologie est particulièrement utilisée en explicabilité car elle permet de justifier l'utilisation de différentes techniques pour expliquer l'un ou l'autre des deux types de systèmes.

Parmi les **explications locales**, on trouve les explications contrastives et contrefactuelles [12]. Ces explications portent sur des décisions individuelles du système : les premières se concentrent sur les différences entre deux résultats possibles pour identifier pourquoi un certain résultat a été préféré à un autre, les secondes explorent des scénarios hypothétiques en identifiant les changements minimaux nécessaires dans les caractéristiques d'entrée pour que le modèle produise une décision différente. Les explications locales sont utilisées à la fois pour les systèmes boîte blanche et boîte noire, même si elles sont aujourd'hui davantage mises en avant pour expliquer les boîtes noires. Elles sont complétées par d'autres dédiées aux boîtes noires qui visent à expliquer les paramètres, les entrées ou les parties du modèle qui ont le plus d'influence sur les décisions [9,14].

Les **explications globales** sur les systèmes boîtes blanches sont usuellement considérées comme triviales par la littérature à cause de la transparence du fonctionnement du moteur d'IA. Cependant, dans de nombreux cas, les informations relatives au fonctionnement de ces systèmes sont nombreuses et ces explications deviennent alors plus complexes à rendre compréhensibles [16].

Ainsi, les deux approches (locales ou globales) présentent un intérêt pour tous les modèles, quelle que soit leur nature (boîtes blanches ou noires). La différence de traitement entre les IA boîtes blanches et les IA boîtes noires en explicabilité réside alors plus dans les techniques de construction de ces explications que dans les approches à utiliser.

2.2 Explicabilité et éducation

Rendre les systèmes d'IA explicables en éducation permet de répondre à plusieurs problématiques [13] : le manque de transparence des systèmes qui peut entraver la confiance des utilisateurs, l'interprétabilité limitée qui peut entraîner une incompréhension de la part des utilisateurs sur les décisions du système et la difficulté d'adaptation du système aux changements d'objectifs ou de situations d'apprentissage.

La mise en place d'explications dans le domaine de l'éducation doit également répondre à des enjeux. En effet, pour qu'elles aient un réel impact sur les utilisateurs du système, elles doivent être comprises par ces derniers. Pour cela, les explications doivent être adaptées aux utilisateurs. Cette adaptation est d'autant plus pertinente que les systèmes éducatifs sont destinés à être utilisés par plusieurs types d'utilisateurs comme les apprenant.e.s et les enseignant.e.s. Ainsi, les arguments attendus dans une explication pour augmenter la confiance et la compréhension seront différents en fonction du rôle de l'utilisateur [11].

Un autre enjeu important à considérer dans la conception des explications est la prise en compte de la charge cognitive engendrée par la compréhension des explications par les utilisateurs [5]. Ainsi, chercher à minimiser cette charge dans la manière de délivrer les explications est crucial pour augmenter leur probabilité d'utilisation par les utilisateurs. Dans ce cadre-là, il a été montré que l'utilisation d'explications visuelles (sous forme de flèches) avait un effet positif sur l'apprentissage et sur la charge cognitive liée à leur utilisation [8].

3 Contexte : le projet ComPer

3.1 Référentiel et profils de compétences

Un des objectifs du projet ComPer était de permettre aux enseignant.e.s de définir des référentiels de compétences qui puissent être opérationnalisables dans des outils informatiques. Ces référentiels peuvent représenter des compétences à plusieurs niveaux de granularité, ainsi que les liens existants entre elles. Ces référentiels sont ensuite exploités dans l'ensemble des outils ComPer, en particulier pour la création de profils de compétences des apprenant.e.s et pour les recommandations des ressources pédagogiques. Un méta-modèle de référentiel impose les contraintes communes à respecter dans leur composition et leur structuration. Il est suffisamment générique pour décrire des référentiels dans plusieurs disciplines éducatives. La figure 1 présente un exemple de profil d'apprenant.e fondé sur un référentiel de compétences à acquérir pour maîtriser les bases opératoires en mathématiques. Cet exemple de référentiel assez simple sera repris tout au long de l'article. Cependant, la contribution proposée a été testée et utilisée sur d'autres référentiels exploités en situation écologique [3].

[3] Les référentiels de compétences sont consultables ici `https://compervm.liris.cnrs.fr/cfe/`

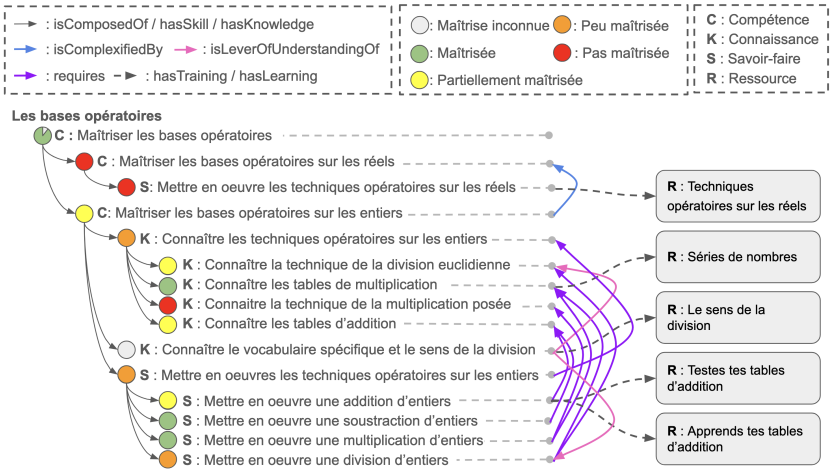


Fig. 1. Référentiel exemple sur les bases opératoires en mathématiques.

Le méta-modèle du projet ComPer définit quatre types d'éléments dans un référentiel [1] : les compétences, les connaissances, les savoir-faire et les ressources pédagogiques (exercices, cours, *etc.*). Les trois premiers éléments seront regroupés sous le terme de *notion* par la suite. Les notions du référentiel sont liées entre elles par des liens sémantiques que l'on peut classer en trois catégories : les liens descendants (*isComposedOf*, *hasSkill*, *hasKnowledge*) en noir à gauche sur la figure 1, les liens transversaux (*requires*, *isComplexifiedBy*, *isLeverOfUnderstandingOf*) en lignes pleines à droite, et les liens avec les ressources (*hasTraining*, *hasLearning*) en pointillé noir à droite. Dans la suite, les liens transversaux seront masqués pour améliorer la lisibilité du profil.

Dans le projet ComPer, les profils de compétences, qui servent à représenter l'avancement d'un.e apprenant.e dans l'acquisition des notions d'un référentiel, sont construits en attribuant un niveau de maîtrise à ces notions [10]. Ces niveaux de maîtrise sont calculés à partir des traces d'activité des apprenant.e.s (notes entre 0 et 1 obtenues aux différentes activités réalisées) et sont représentés sur le profil par des pastilles de couleur.

Par exemple, dans le profil de la figure 1, la compétence "Maîtriser les bases opératoires sur les entiers“ est *partiellement maîtrisée* et se compose de deux connaissances et d'un savoir-faire. La première connaissance, "Connaître les techniques opératoires sur les entiers“, est un prérequis du savoir-faire "Mettre en œuvre les techniques opératoires sur les entiers“.

3.2 Moteur de recommandations

Dans le projet ComPer, la personnalisation de l'apprentissage se traduit par la recommandation de ressources. Pour obtenir une recommandation, les enseignant.e.s et/ou les apprenant.e.s doivent définir un ou plusieurs **objectifs**

pédagogiques en déclarant quelles **notions du référentiel** iels souhaitent travailler et avec quelle **intention pédagogique**. Les intentions pédagogiques permettent de préciser la manière dont les utilisateur.ice.s désirent travailler la notion. L'intention a un impact important sur les résultats du système de recommandation. Il existe 6 intentions dans ComPer : prérequis, découverte, soutien, perfectionnement, révision et levier de compréhension [15]. Ensuite, le système exploite à la fois la **structure du référentiel** de compétences et les **valeurs du profil** d'un.e apprenant.e pour construire une liste ordonnée et personnalisée de ressources pédagogiques que l'apprenant.e devrait travailler pour atteindre les objectifs pédagogiques définis.

L'algorithme permettant d'obtenir ces recommandations fonctionne selon une approche symbolique (IA de type boite blanche), avec un parcours de graphe du référentiel de compétences. Le parcours est guidé par des règles élaborées par des expert.e.s et modifiables par les enseignant.e.s *via* une interface de paramétrage. Un ensemble de paramètres fixé est appelé **stratégie pédagogique**. Par exemple, dans la règle exposée dans la figure 2, qui définit comment parcourir le référentiel de compétences à partir d'un objectif pédagogique d'intention *révision*, les paramètres fixés dans la stratégie pédagogique par défaut sont notés en orange. Le système exploite également un certain nombre de poids modifiables, qui permettent d'ajuster le comportement de l'algorithme en fonction des intentions. L'algorithme se décompose en trois grandes étapes : sélection des notions, ordonnancement des notions et choix des ressources. Ces trois étapes s'effectuent pour chaque objectif pédagogique fourni en paramètre de la demande de recommandations [15].

La **sélection des notions** s'appuie sur le parcours du profil de l'apprenant.e pour récupérer les notions sur lesquelles l'apprenant.e aurait besoin de travailler. Pour cela, l'algorithme part de la notion objectif et utilise des règles similaires à celles de la figure 2. En fonction de l'intention, la notion subit une phase de sélection dans laquelle elle peut être sélectionnée comme notion à travailler et/ou ajoutée à la liste des objectifs pédagogiques avec une nouvelle intention. Ensuite, l'algorithme va parcourir les liens descendants à partir de cette notion pour récupérer leurs notions cibles. En fonction de l'intention, ces dernières peuvent soit passer en phase de sélection, soit l'algorithme continue son parcours à partir de ces notions mais en parcourant cette fois-ci les liens transversaux. Les notions transversales ainsi récupérées font ensuite l'objet du traitement de sélection. Ainsi, à partir d'un même objectif pédagogique de départ, plusieurs types de liens peuvent être parcourus et plusieurs objectifs pédagogiques intermédiaires peuvent être créés. Cela peut aboutir à la sélection multiple d'une même notion par des parcours du référentiel différents, appelés **chemins** par la suite.

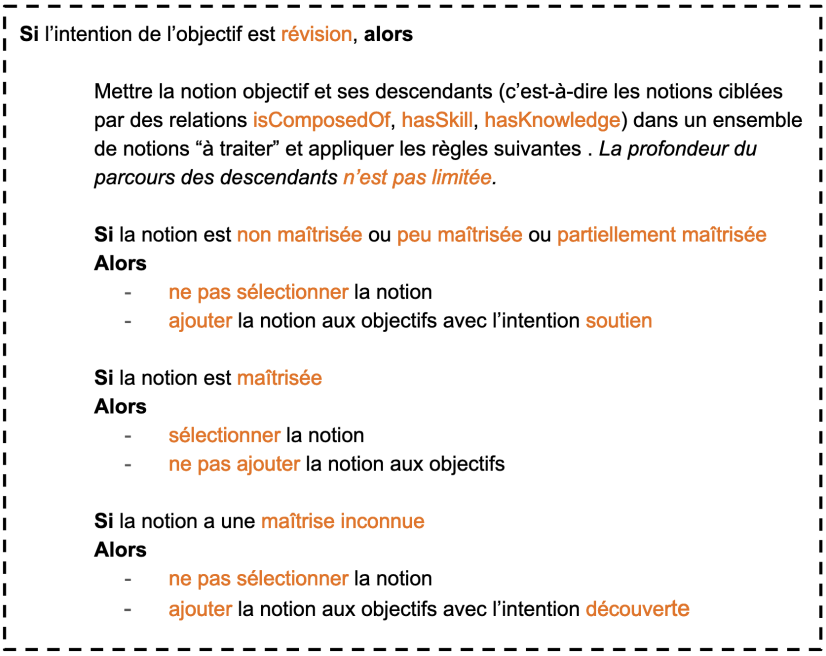


Fig. 2. Règle pour l'intention *révision*.

L'**ordonnancement des notions** utilise un poids de priorité calculé lors de la sélection des notions à travailler. Ce dernier dépend de l'intention de l'objectif pédagogique et des liens parcourus pour atteindre la notion considérée. Dans le cas où la notion est sélectionnée plusieurs fois, les poids des différents parcours sont agrégés avec une pondération dépendante de l'intention de départ. Le **choix des ressources** permet de récupérer et de filtrer, selon le paramétrage, les ressources rattachées aux notions sélectionnées. Ce sont ces ressources qui seront ensuite recommandées à l'apprenant.e.

Tous les détails de l'exécution des trois étapes de l'algorithme sont tracés dans des fichiers. Ces derniers sont cruciaux pour la suite et serviront à la création d'explications.

4 Contributions

Notre question de recherche est : **Comment assister les enseignants dans le paramétrage d'un système de recommandations pédagogiques ?**

Comme nous l'avons indiqué précédemment, le système de recommandations de ComPer trace toutes les étapes de son raisonnement, ainsi que les différents chemins suivis lors du parcours du référentiel. **Notre approche consiste à exploiter ces traces comme sources de la construction d'explications** devant facilier le paramétrage du système de recommandations.

Notre construction d'explications repose sur 3 étapes présentées dans la figure 3 : le pré-traitement des traces et l'identification de motifs fréquents, puis la génération à proprement parler des explications. Ces étapes sont respectivement détaillées dans les sections 4.1 et 4.2. Les choix de présentation des explications générées aux enseignant.e.s seront présentés dans la section 4.3.

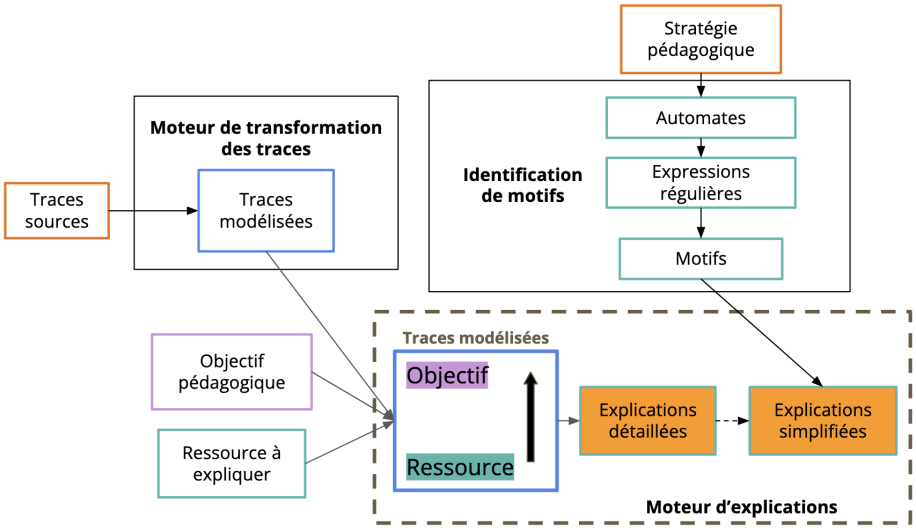


Fig. 3. Démarche utilisée pour répondre à la problématique.

4.1 Pré-traitement des traces et identification de motifs fréquents

Pour faciliter leur exploitation, il était nécessaire de réaliser un travail de restructuration des fichiers de traces. Ces derniers sont généralement longs et ne contiennent que des informations de bas niveau pas toujours pertinentes pour la construction de nos explications. Ainsi, nous avons utilisé une approche de traces modélisées [3] pour réaliser cette tâche de transformation des fichiers. Cela nous a alors permis d'abstraire les informations relatives à la phase de sélection des notions dans les traces en représentant le parcours par plusieurs événements appelés **traitements d'une notion** (cf. figure 4). Ces traitements peuvent être une sélection en tant que notion à travailler. Dans ce cas, la raison de la sélection de la notion est précisée. Sinon, il s'agit d'un ajout de la notion aux objectifs pédagogiques et l'intention associée à l'objectif créé est donnée.

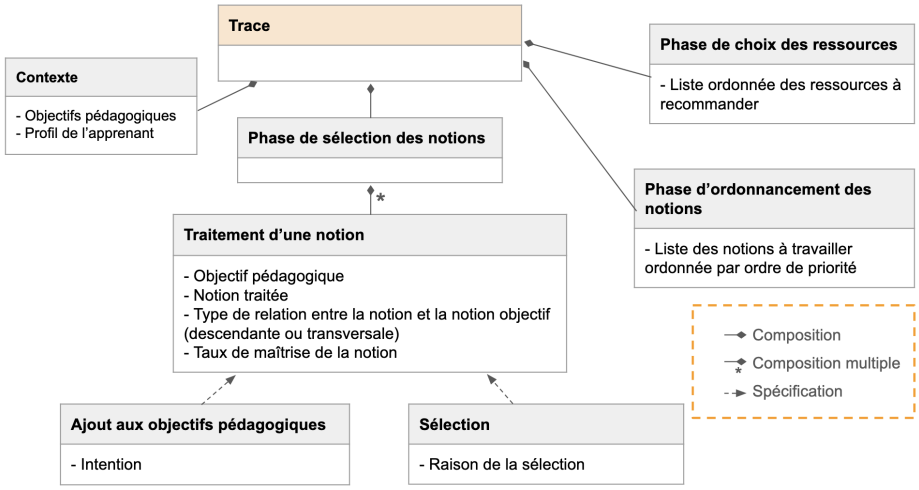


Fig. 4. Contenu des traces après le pré-traitement.

Le travail sur les traces a permis de mettre en évidence l'existence d'une grande diversité dans les suites de liens parcourus dans les référentiels en termes de nature des liens suivis. Cela a alors motivé la volonté d'identifier des motifs fréquents parmi ces derniers.

Pour identifier les motifs fréquents dans les chemins possibles à partir d'une stratégie pédagogique, nous avons choisi de représenter l'ensemble de ses règles contenues dans la stratégie par défaut sous forme d'automates. Cela nous a alors permis d'identifier sous forme d'expressions régulières tous les chemins possibles à partir de toutes les intentions pédagogiques possibles depuis un objectif de départ. Il était ensuite possible d'extraire les motifs fréquents dans les chemins en cherchant les sous-parties d'expressions régulières se retrouvant dans plusieurs expressions régulières ainsi construites.

Nous avons pu identifier quatre motifs qui reviennent plusieurs fois pour chaque intention et/ou qui se retrouvent dans plusieurs traces : suivi d'un ou plusieurs liens descendants, suivi d'un ou plusieurs liens transversaux, suivi de liens de pré-requis puis de liens descendants, et suivi de liens descendants puis de pré-requis. Les relations de descendance et de pré-requis étant transitives, ces motifs ont été renommés en : exploration de descendants (P1), exploration de transversaux (P2), exploration de descendants de pré-requis (P3) et exploration de pré-requis de descendants (P4).

4.2 Génération d'explications de deux niveaux de granularité

Le travail sur les traces nous a permis de construire un premier type d'explication, les explications détaillées, que l'identification de motifs nous a permis de simplifier pour former un second type d'explication, les explications simplifiées.

D'un côté, les **explications détaillées** ont pour objectif d'exposer le parcours réellement effectué par l'algorithme, en utilisant les traces pour retrouver les séquences d'activation de règles et donc les liens parcourus pour atteindre une ressource recommandée. Pour cela, le moteur prend en entrée les traces modélisées, une ressource recommandée à expliquer et l'objectif pédagogique associé. Cela permet alors de retrouver dans les traces l'ensemble des parcours du référentiel menant à la ressource, en remontant le fichier de traces depuis la sélection de la notion parent de la ressource jusqu'au traitement de l'objectif pédagogique fourni en entrée du moteur. L'objectif est de créer une structure de données sous forme de liste contenant les différents chemins retrouvés dans cette exploitation des traces. Cette structure contient alors pour chaque parcours les détails relatifs à la sélection de la notion parent de la ressource par l'algorithme (règle activée, objectif pédagogique traité, type de relation suivie). Si des objectifs pédagogiques intermédiaires ont été utilisés dans ce parcours, leur création est également stockée avec leurs informations associées (règle activée, objectif pédagogique traité, intention associée et type de relation suivie).

D'un autre côté, les **explications simplifiées** ont pour objectif d'utiliser les motifs identifiés dans le travail théorique sur les règles afin de permettre l'exposition des parcours du référentiel de façon plus globale et donc plus simple à comprendre pour les enseignant.e.s. Cela se fait en retrouvant ces motifs dans les explications détaillées. Les étapes du parcours correspondant à un motif sont alors fusionnées, ce qui permet au moteur de fournir en sortie une liste des motifs présents dans les chemins des traces avec les notions qui relient ces motifs.

4.3 Explications des recommandations à l'enseignant.e

Le premier but des explications est de faciliter le paramétrage du système de recommandations par les enseignant.e.s, en augmentant leur compréhension du fonctionnement du système. Ainsi, il semblait pertinent de délivrer ces explications aux enseignant.e.s au sein même de l'interface de paramétrage existante.

Présentation des explications simplifiées. Le but des explications simplifiées est de montrer à l'enseignant.e les parcours du référentiel de compétences pour atteindre la ressource recommandée à expliquer, sans pour autant entrer dans le détail de l'activation des différentes règles. Ainsi, il a été choisi de les délivrer principalement sous forme visuelle, en dessinant des flèches sur le profil affiché (haut de la figure 5). Nous avons choisi un format visuel en superposition du profil, que les enseignant.e.s sont habitué.e.s à voir, dans le but de réduire la charge cognitive de l'utilisateur [8]. La couleur de ces flèches correspond à la nature des chemins qu'elles représentent. Si nous reprenons la définition des motifs de la section 4.1, une flèche blanche représente le suivi des liens descendants (motif P1), une flèche violette représente le suivi des liens transversaux (motif P2), et une flèche verte correspond aux motifs P3 et P4. Afin de ne pas surcharger l'affichage, lorsque l'algorithme prend plusieurs chemins de natures

différentes entre deux mêmes notions, les flèches représentant ces différents chemins sont fusionnées et représentées par une flèche bleue.

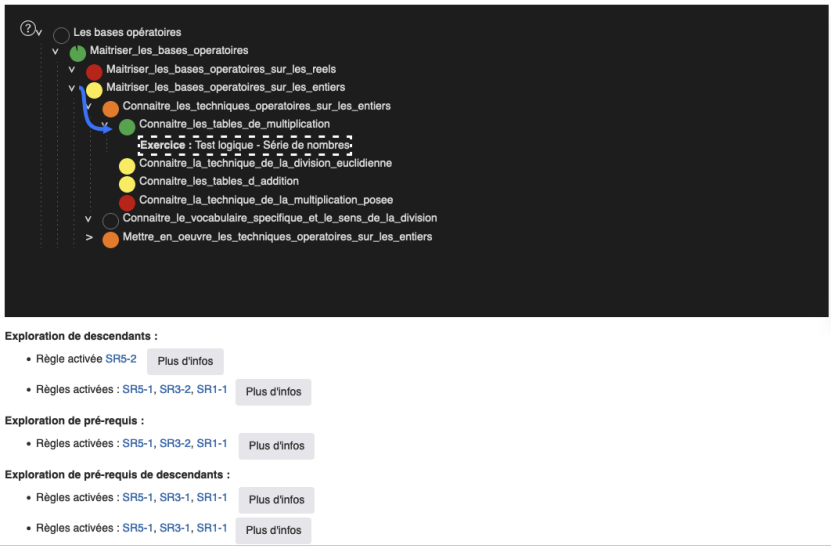


Fig. 5. Exemple d'affichage d'une explication simplifiée.

Dans l'exemple présenté en figure 5, l'utilisateur a demandé des explications sur la recommandation "Test logique - Série de nombres" obtenue à partir de l'objectif pédagogique "Maîtriser les bases opératoires sur les entiers" avec l'intention *révision*. Pour atteindre la notion parent de cette ressource à partir de l'objectif, l'algorithme a emprunté 5 chemins avec 3 motifs différents. C'est donc pour cette raison que la flèche dessinée est bleue. De plus, pour faciliter la lecture de ces explications, la ressource expliquée est mise en valeur. En outre, les éléments du profil qui n'apparaissent pas dans le parcours du référentiel sont réduits (repliés) au maximum.

La bascule entre les deux types d'explications apparaît lorsque l'utilisateur passe sa souris sur une flèche. La liste des différents chemins qu'elle représente apparaît alors sous le profil (bas de la figure 5). Il est ainsi possible de demander plus d'informations pour un des chemins, ce qui fait apparaître l'explication détaillée correspondante.

Présentation des explications détaillées. L'objectif de ces explications détaillées est de faire comprendre à l'enseignant.e le lien entre le paramétrage du système et le parcours du référentiel par un certain chemin. Pour rappel, un chemin correspond à l'activation successive des règles de la stratégie pédagogique.

Ainsi, pour expliquer l'apparition d'un chemin, les raisons et les conséquences de l'activation de chaque règle utilisée dans ce dernier sont montrées à l'enseignant.e *via* des flèches sur le profil (figure 6). Un texte expliquant ce que représente la flèche est accessible lorsque l'utilisateur passe sa souris sur une flèche. Comme pour les explications simplifiées, la couleur des flèches représente la nature du lien entre les deux notions qu'elle relie. Une flèche blanche représente un lien descendant tandis qu'une flèche violette représente un lien transversal.

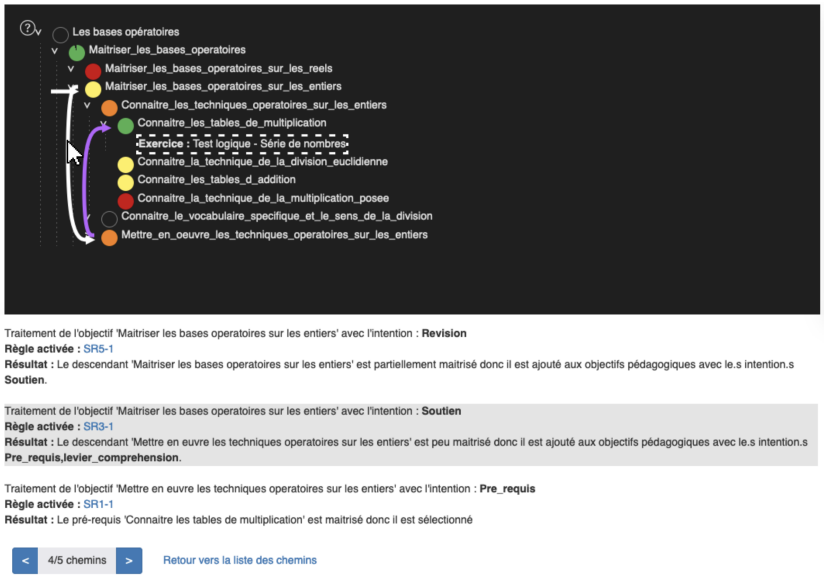


Fig. 6. Exemple d'affichage d'une explication détaillée.

La figure 6 représente l'explication détaillée associée au quatrième chemin montré dans l'exemple d'explication simplifiée de la figure 5 : le premier chemin d'exploration de prérequis de descendants. Ici, on voit alors bien que l'algorithme suit d'abord deux liens descendants puis un lien transversal. Des explications textuelles détaillant ce que représentent les flèches sont mises en avant quand on passe la souris sur celles-ci. Dans l'exemple, la souris est sur la grande flèche blanche. On peut observer que lors du traitement de l'objectif pédagogique intermédiaire "Maîtriser les bases opératoires sur les entiers" avec l'intention soutien, la règle *SR3-1* a été activée. L'activation de cette règle a alors eu pour conséquence d'ajouter aux objectifs pédagogiques la notion descendante "Mettre en oeuvre les techniques opératoires sur les entiers" avec les intentions levier de compréhension et pré-requis. Le taux de maîtrise de la notion est également donné, car c'est la raison de son ajout.

5 Expérimentations

Nous avons mis à l'essai nos propositions afin d'obtenir un retour qualitatif sur deux points. Le premier concernait l'utilisabilité de l'interface : est-ce qu'elle permet aux enseignant.e.s d'atteindre les explications d'un point de vue ergonomique ? Le second concernait l'utilité des explications : est-ce que les explications permettent aux enseignant.e.s de mieux comprendre le fonctionnement du système de recommandations et d'ainsi faciliter son paramétrage ? Ici, nous considérons un paramétrage comme étant réussi, si ce dernier assure que le système se comporte en cohérence avec les souhaits pédagogiques de l'enseignant.e qui l'utilise.

5.1 Population

Nous avons rencontré deux enseignantes d'équipes pédagogiques différentes. La première est enseignante en Sciences des technologies de l'information et de la communication (STIC) et a participé au projet ComPer. La seconde est enseignante en physique-chimie au lycée et utilise les outils ComPer sur un référentiel de compétences défini en accord avec ses collègues [4]. Ainsi, toutes deux connaissent les principes de l'APC, le méta-modèle du référentiel de compétences et les principes du fonctionnement du moteur de recommandations (choix d'une notion avec une intention pédagogique, exploitation du profil de l'apprenant.e, *etc.*).

5.2 Tâche réalisée

Les enseignantes ont manipulé l'interface sur le référentiel utilisé en exemple dans ce document. Elles étaient guidées sur 4 cas d'utilisation que nous avions identifié au préalable. Un cas d'utilisation est une demande d'explication sur une certaine ressource recommandée à partir d'un objectif et d'un profil de compétences définis. Aussi, nous avions construit 4 profils d'apprenant.e.s fictifs. Chacun d'eux permet de représenter des avancements dans l'acquisition des compétences : compétences toutes travaillées ou seulement pour une partie du référentiel, compétences toutes maîtrisées ou seulement certaines d'entre elles ou aucune, *etc.* Chaque cas permettait de représenter différentes situations menant à des stratégies d'explications et donc à des motifs de parcours de référentiel différents : parcours d'un seul chemin avec uniquement des liens descendants, parcours d'un seul chemin avec suivi de plusieurs types de liens, parcours de plusieurs chemins, comparaison de deux résultats différents de recommandation sur deux profils différents avec le même objectif pédagogique. Des questions leur étaient posées tout au long de leur utilisation du système. L'ensemble de leurs réponses et retours était pris en note.

[4] L'enseignante et ses collègues font partie du groupe PEGASE `https://pegase.ens-lyon.fr/`

5.3 Premiers éléments sur l'utilisabilité

Cette évaluation s'est faite *via* l'observation de l'utilisation du système par les enseignantes. Des questions leur étaient également posées pour évaluer leur compréhension des différents éléments de l'interface. Ainsi, la capacité des enseignantes à utiliser l'interface et à répondre aux questions sans guidage extérieur de notre part nous fournit un éclairage sur l'utilisabilité de notre outil. Leurs réponses orales aux questions et leurs interactions avec le système étaient prises en note. Ici, il est important de souligner que l'objectif de cette évaluation est de donner un éclairage sur l'utilisabilité et non de l'affirmer ou de l'infirmer. Le protocole mis en place présente trop de biais pour cela : le manque de standardisation de critères observables de mesure, notre rôle actif dans le protocole, la non-définition en amont de l'utilisabilité, la trop petite taille de la population.

Les deux expérimentations ont montré un problème en termes d'utilisabilité. En effet, les enseignantes ont eu beaucoup de mal à prendre en main l'interface et ont eu besoin de guidage à de nombreuses reprises pour répondre aux questions. Par exemple, elles ont souligné le manque d'une légende pour les couleurs des flèches, qui empêchait la compréhension des explications, mais aussi la difficulté de compréhension d'éléments permettant de faire le lien entre les explications simplifiées et les explications détaillées.

5.4 Premiers éléments sur l'utilité

Les explications ayant pour objectif d'améliorer la compréhension du fonctionnement du système de recommandations et de faciliter son paramétrage, il était nécessaire d'évaluer ces deux points pour juger de leur utilité. Pour le premier point, il était demandé aux enseignantes de chercher pourquoi une ressource était recommandée pour un apprenant.e donné alors qu'elle ne l'était pas pour un autre. Pendant les expérimentations, les enseignantes ont réussi à trouver la raison de la différence de recommandations en utilisant les explications. Cela tend à indiquer que celles-ci aident à comprendre le fonctionnement du système. Pour le deuxième point d'évaluation de l'utilité, les enseignantes devaient réaliser une tâche de paramétrage. Elles avaient alors pour consigne de faire apparaître une recommandation dans un cas où le système ne recommandait aucune ressource. Durant les expérimentations, elles se sont aidées des explications pour trouver un paramétrage du système répondant à leur besoin. Cela peut signifier que les explications ont un impact sur la compréhension du lien entre paramétrage et recommandations.

6 Conclusion et limites

Dans cet article, nous avons présenté un outil qui a pour objectif d'assister les enseigant.e.s dans le paramétrage d'un moteur de recommandations pédagogiques. Pour cela, nous avons choisi d'utiliser une approche permettant d'expliquer aux enseigant.e.s les raisons pour lesquelles ce moteur de recommandations choisit de fournir certaines ressources pédagogiques à certains apprenant.e.s. Les résultats des premières mises à l'essai menées semblent indiquer que la contribution tend à aider les enseignant.e.s à mieux comprendre le fonctionnement du système de recommandations et à paramétrer le système, malgré une difficulté à prendre en main l'interface proposée à cause d'un manque général d'utilisabilité.

Cependant, ces résultats sont à confirmer car les expérimentations menées pour les obtenir présentent des limites. Dans un premier temps, il est important de souligner la petite taille de la population (2 enseignantes), qui ne permet pas de penser que les résultats obtenus sont généralisables. De plus, pour éviter d'avoir à introduire les outils ComPer et les notions relatives à l'APC, nous avons choisi de réaliser les mise à l'essai avec des membres d'équipes pédagogiques connaissant déjà le projet. Il est également possible de critiquer l'absence d'un groupe contrôle qui testerait l'outil de paramétrage de ComPer sans explications. En effet, cela nous empêche d'affirmer avec certitude que les résultats obtenus sont dûs aux explications. Ainsi, une perspective est de poursuivre l'évaluation des explications avec une population plus grande et plus diverse et avec un groupe contrôle. De plus, afin de réduire les biais dans l'évaluation de l'utilisabilité, le protocole mis en place devra être revu en utilisant un test normalisé existant dans la littérature comme le *System Usability Scale* [2] ou le *Technology Acceptance Model* [4].

Plus globalement, le travail présenté est un premier pas vers la mise en place d'une évolution continue du système de recommandations en collaboration humain-machine. Cependant, beaucoup de choses restent à faire. Par exemple, il peut être intéressant de mettre en place des explications dans les autres outils ComPer, afin de favoriser l'appropriation des systèmes d'IA. De plus, dans l'objectif d'améliorer l'ensemble des outils ComPer, il va également être nécessaire de s'intéresser aux potentiels problèmes d'efficacité et de biais algorithmiques dans les systèmes d'IA de ComPer. Pour cela, il faudra mettre en place un autodiagnostic de ces problèmes afin de pouvoir les détecter automatiquement et ensuite les corriger, pour l'amélioration de ces systèmes d'IA.

Remerciements

Ce travail a été soutenu par les projets ComPer (ANR-18-CE38-0012) et TeaTIME (ANR-23-CE38-0012) financés par l'ANR.

Références

1. Broisin, J., Bey, A. : Comper lot 1 : Méta-modèle du référentiel de compétences fondé sur le web sémantique (Mar 2020), `https://comper.fr/sites/`

default/files/2020-12/L1.1%20Me%CC%81ta-mode%CC%80le%20re%CC%81fe%CC%81rentiels_2.pdf
2. Brooke, J., et al. : Sus-a quick and dirty usability scale. Usability evaluation in industry **189**(194), 4–7 (1996)
3. Cordier, A., Lefevre, M., Champin, P.A., Georgeon, O., Mille, A. : Trace-Based Reasoning - Modeling interaction traces for reasoning on experiences. In : McCarthy, P. (ed.) The 26th International FLAIRS Conference. pp. 1–15. St. Pete Beach, Florida, United States (May 2013), https://hal.science/hal-00830444
4. Davis, F.D. : A technology acceptance model for empirically testing new end-user information systems : Theory and results. Ph.D. thesis, Massachusetts Institute of Technology (1985)
5. Donoso-Guzmán, I., Ooge, J., Parra, D., Verbert, K. : Towards a Comprehensive Human-Centred Evaluation Framework for Explainable AI (Jul 2023), arXiv :2308.06274 [cs]
6. Ghezir, S., Naimie, Z., Leng, C.H., Shagholi, R., Abuzaid, R.A. : Review study : Competency-based approach implementation in educational system. New Trends and Issues Proceedings on Humanities and Social Sciences **8**(3), 21–29 (Oct 2021). https://doi.org/10.18844/prosoc.v8i3.6156, https://un-pub.eu/ojs/index.php/pntsbs/article/view/6156, number : 3
7. Hassija, V., Chamola, V., Mahapatra, A., Singal, A., Goel, D., Huang, K., Scardapane, S., Spinelli, I., Mahmud, M., Hussain, A. : Interpreting Black-Box Models : A Review on Explainable Artificial Intelligence. Cognitive Computation **16**(1), 45–74 (Jan 2024). https://doi.org/10.1007/s12559-023-10179-8
8. Jamet, E., Fernandez, J. : Enhancing interactive tutorial effectiveness through visual cueing. Educational Technology Research and Development **64**(4), 631–641 (Aug 2016). https://doi.org/10.1007/s11423-016-9437-6
9. Lundberg, S.M., Lee, S.I. : A Unified Approach to Interpreting Model Predictions. In : Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (eds.) Advances in Neural Information Processing Systems. vol. 30. Curran Associates, Inc. (2017), https://proceedings.neurips.cc/paper_files/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf
10. Mandin, S., Guin, N. : Basing learner modelling on an ontology of knowledge and skills. In : Sampson, D.G., Spector, J.M., Chen, N.S., Huang, R., Kinshuk (eds.) IEEE International Conference on Advanced Learning Technologies. pp. 321–323. IEEE Computer Society, Athènes, Greece (Jul 2014), https://hal.science/hal-01301064
11. Millecamp, M., Htun, N.N., Conati, C., Verbert, K. : To explain or not to explain : the effects of personal characteristics when explaining music recommendations. In : Proceedings of the 24th International Conference on Intelligent User Interfaces. pp. 397–407. IUI '19, Association for Computing Machinery, New York, NY, USA (Mar 2019). https://doi.org/10.1145/3301275.3302313, https://doi.org/10.1145/3301275.3302313
12. Miller, T. : Explanation in artificial intelligence : Insights from the social sciences. Artificial Intelligence **267**, 1–38 (Feb 2019). https://doi.org/10.1016/j.artint.2018.07.007
13. Rachha, A., Seyam, M. : Explainable AI In Education : Current Trends, Challenges, And Opportunities (Apr 2023). https://doi.org/10.1109/SoutheastCon51012.2023.10115140, pages : 239
14. Ribeiro, M.T., Singh, S., Guestrin, C. : "Why Should I Trust You ?" : Explaining the Predictions of Any Classifier (Aug 2016). https://doi.org/10.48550/arXiv.1602.04938, http://arxiv.org/abs/1602.04938, arXiv :1602.04938 [cs]

15. Sablayrolles, L., Lefevre, M., Guin, N., Broisin, J. : Design and Evaluation of a Competency-Based Recommendation Process. In : Intelligent Tutoring Systems : 18th International Conference, ITS 2022, Bucharest, Romania, June 29 – July 1, 2022, Proceedings. pp. 148–160. Springer-Verlag, Berlin, Heidelberg (Jun 2022). https://doi.org/10.1007/978-3-031-09680-8_14
16. Saleem, R., Yuan, B., Kurugollu, F., Anjum, A., Liu, L. : Explaining deep neural networks : A survey on the global interpretation methods. Neurocomputing **513**, 165–180 (Nov 2022). https://doi.org/10.1016/j.neucom.2022.09.129
17. Schwalbe, G., Finzel, B. : A Comprehensive Taxonomy for Explainable Artificial Intelligence : A Systematic Survey of Surveys on Methods and Concepts. Data Mining and Knowledge Discovery (Jan 2023). https://doi.org/10.1007/s10618-022-00867-8, arXiv :2105.07190 [cs]
18. Tiwari, R. : Explainable AI (XAI) and its Applications in Building Trust and Understanding in AI Decision Making. International journal of scientific research in engineering and management **07**(01) (Jan 2023). https://doi.org/10.55041/IJSREM17592

Apprentissages autorégulés et expérience dans un contexte de formation hybride universitaire

Enosch Guelaybe Djiezion[1]

[1]INSPE Nice, Université Côte d'Azur, LINE, UPR CA02, Nice, France
Enosch.Djiezion@univ-cotedazur.fr

Résumé. Cet article traite de la relation entre les apprentissages autorégulés (AAR) et l'expérience du contexte de formation hybride à l'Université Paris Cité. L'étude, motivée par l'hybridation de la formation due à l'augmentation du nombre d'étudiants et les défis logistiques, explore la corrélation entre cette variable et l'expérience des étudiants en contexte de formation hybride. En utilisant une approche quantitative, nous avons analysé les réponses de 110 étudiants, concluant que l'AAR évolue en fonction de l'expérience des étudiants, suggérant des implications importantes pour l'optimisation des stratégies pédagogiques dans des environnements hybrides.

Mots-clés : Aprentissages autorégulés, formation hybride, échelle des apprentissages autorégulés en ligne (EAREL), Coefficient de corrélation

Abstract. This article examines the relationship between self-regulated learning (SRL) and the experience of hybrid learning environments at Université Paris Cité. The study, driven by the increasing hybridization of education due to growing student numbers and logistical challenges, explores the correlation between SRL and students' experiences in hybrid training contexts. Using a quantitative approach, we analyzed the responses of 110 students, concluding that SRL evolves ac- cording to students' experiences, suggesting important implications for optimizing pedagogical strategies in hybrid learning environments.

Keywords: Self-regulated learning, hybrid learning, Online Self-Regulated Learning Scale (OSRLS), correlation coefficient

1 Introduction

Les apprentissages autorégulés (AAR) est un concept essentiel pour comprendre et améliorer la réussite scolaire des étudiants. Il s'agit d'un processus où les étudiants gèrent de manière autonome leurs pensées, leurs émotions et leurs comportements pour atteindre leurs objectifs personnels (Zimmerman et Schunk, 2011), dans le cadre de l'acquisition de la connaissance.

La transition récente de la licence en sciences de l'éducation de l'Université Paris Cité, qui intègre désormais des cours dès la première année, a conduit à une augmentation importante du nombre d'étudiants. Cette expansion a mis en évidence des défis logistiques et pédagogiques, incitant à l'adoption de formations hybrides combinant enseignement présentiel et à distance. Cette modalité vise à tirer parti des avantages de chaque approche pour offrir un apprentissage optimal.

Cependant, le passage à une formation hybride soulève des questions sur les apprentissages des étudiants dans un tel contexte, notamment sur les interactions entre l'expérience du contexte d'apprentissage en formation hybride et les AAR des étudiants. Bien que des recherches antérieures aient montré que ces deux variables sont liées à la réussite scolaire, leur relation spécifique dans un environnement hybride reste peu explorée. Cette étude vise donc à examiner la corrélation les AAR des étudiants et leur expérience d'apprentissage dans un contexte de formation hybride à l'Université Paris Cité, en se basant sur des hypothèses soutenues par des recherches antérieures.

En s'appuyant sur les théories de Zimmerman (2002) et d'autres chercheurs, la revue de la littérature effectuée explore les fondements des AAR, ainsi que leur pertinence dans le cadre d'une formation hybride. La problématique centrale est de déterminer comment cette variable évolue en fonction de l'expérience des étudiants dans un environnement d'apprentissage mixte, en posant l'hypothèse que les AAR évoluent en fonction de l'expérience de l'étudiant du dispositif universitaire d'apprentissage mixte. Les résultats de cette étude fourniront des données précieuses pour améliorer les stratégies pédagogiques et soutenir la réussite des étudiants dans des contextes hybrides.

Nous commencerons ci-après par poser le cadre théorique de la recherche. Nous exposerons ensuite la problématique de l'étude, puis expliciterons la méthodologie de sa mise en œuvre. Enfin, nous présenterons les résultats de l'étude et leur discussion.

2 Revue de la littérature

2.1 Les apprentissages autorégulés

Zimmerman et Schunk (2011) considèrent l'apprentissage autorégulé comme un processus où les apprenants gèrent et maintiennent de façon autonome leurs pensées, leurs émotions et leurs comportements afin de réaliser leurs objectifs personnels. Paquelin et Choplin (2003) décrivent l'apprentissage autorégulé comme la capacité de l'apprenant à ajuster ses stratégies d'apprentissage en fonction des objectifs qu'il s'est fixés de manière autonome. Ce processus implique une évaluation des stratégies utilisées et, si nécessaire, leur modification pour atteindre les objectifs. Pintrich (2000) partage cette vision, définissant l'apprentissage autorégulé comme un processus actif et constructif

qui permet à l'apprenant de définir des objectifs d'apprentissage et de réguler ses cognitions, ses motivations et ses comportements en fonction de l'environnement. Ainsi, l'apprenant autorégulé planifie et anticipe ses actions en s'appuyant sur ses connaissances et les ressources disponibles, contrôle sa motivation et son attention, et s'autoévalue pour ajuster ses actions si nécessaire (Zimmerman, 2002).

D'après Pintrich (2000), l'apprentissage autorégulé repose sur quatre principes fondamentaux : premièrement, l'apprenant est un acteur actif de son apprentissage, construisant progressivement significations, objectifs et stratégies à partir des informations obtenues de l'environnement et de sa propre structure cognitive. Deuxièmement, l'apprenant évalue, contrôle et ajuste certains aspects de ses pensées, comportements et motivations tout en prenant en compte les caractéristiques de l'environnement. Troisièmement, il évalue son processus d'apprentissage en comparant les objectifs fixés avec les résultats obtenus et, quatrièmement, il ajuste ses comportements et stratégies en conséquence

Selon Cosnefroy (2010), les modèles d'apprentissage autorégulé se structurent autour de trois phases séquentielles :

- Préparation : l'apprenant se prépare à entreprendre les activités d'apprentissage en fixant des objectifs, en planifiant (aspect cognitif) et en évaluant le contexte d'apprentissage en termes de défis et d'opportunités (aspect motivationnel).
- Engagement : durant cette phase centrale, l'apprenant s'implique activement dans la tâche, appliquant des stratégies d'autorégulation pour atteindre les objectifs fixés.
- Évaluation : l'apprenant évalue son activité d'apprentissage, réorganise ses connaissances et ajuste ses croyances et perceptions en fonction des succès et des échecs, ainsi que des performances réalisées.

2.2 La formation hybride

Une formation hybride (figure 1), également désignée par « formation mixte » ou blended learning, est un mode de formation qui combine l'e-learning et l'enseignement présentiel au sein d'un même parcours pédagogique cohérent.

Selon Charlier et al. (2006), il s'agit d'un mode de formation qui se situe à la croisée de l'enseignement à distance et en présentiel. Il combine et parfois intègre certaines caractéristiques propres à ces deux modalités de formation.

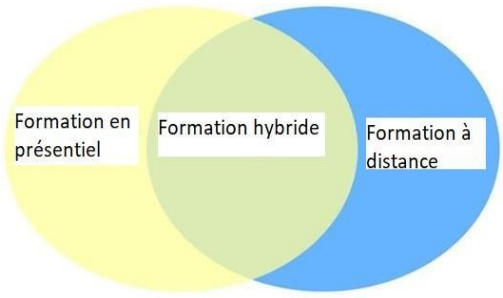


Fig. 1. La formation hybride

La mise en œuvre de modalités mixtes se fait essentiellement dans l'optique de tirer les meilleurs avantages du présentiel et du distanciel pour atteindre l'apprentissage « optimal ». Il combine ainsi les effets positifs de l'enseignement présentiel et ceux de l'enseignement à distance. D'après une étude menée par les chercheurs Means, Toyama, Murphy, Baki et Jones (2009) sur l'enseignement en ligne aux É.-U. de 1996 à 2008, l'apprentissage hybride ou mixte (blended learning) qui combine l'enseignement en ligne avec l'enseignement présentiel a plus d'avantages que l'apprentissage purement présentiel, ou purement en ligne. Ali, A. et al. (2023) et Rayyan, M. et al. (2024) abondent dans le même sens, en soutenant que l'apprentissage hybride améliore significativement l'auto-efficacité des étudiants, ce qui est positivement corrélé à une meilleure performance académique.

On désigne ainsi par le terme de blended learning la formation dispensée selon plusieurs modalités d'apprentissage combinées. Les principales sont :

- Le présentiel, c'est-à-dire entre deux personnes ou plus physiquement présentes au même endroit ;
- Le distanciel asynchrone, c'est-à-dire par l'intermédiaire d'un contenu de formation e- learning ;
- Le distanciel synchrone, c'est-à-dire par le biais d'une visioconférence.

La majorité des programmes d'apprentissage mixte se fait selon l'un des quatre modèles suivants (Garrison et Kanuka, 2004) :

- Rotation : le modèle de rotation est caractérisé par une alternance planifiée entre différentes modalités d'apprentissage, notamment en ligne et en présentiel. Les étudiants suivent un emploi du temps structuré qui leur permet de passer d'une activité à une autre.
- Flex : dans ce modèle, l'e-learning est l'épine dorsale de l'apprentissage, même si, de temps en temps, les étudiants peuvent pratiquer des activités hors ligne.
- À la carte : l'étudiant choisit lui-même les modalités qui lui conviennent. Il peut suivre le cours à la carte, soit sur le campus ou en ligne.
- Virtuel enrichi : le modèle virtuel enrichi repose principalement sur l'apprentissage en ligne, avec quelques sessions en présentiel pour approfondir certains aspects ou réaliser des activités spécifiques. Contrairement au modèle de rotation, la majorité de l'apprentissage se fait à distance et la présence physique en classe est moins fréquente et plus ponctuelle.

3 Problématique, question et hypothèse de recherche

3.1. Problématique et question de recherche

Depuis 2014, la licence en sciences humaines et sociales, mention sciences de l'éducation, à l'Université Paris Cité, est accessible dès la première année (L1). Avant cette date, l'accès n'était possible qu'à partir de la troisième année de licence (L3). Ce changement a conduit à une augmentation importante du nombre d'étudiants, passant d'une cinquantaine à plus de 250 en moyenne. Cette croissance des effectifs a posé des

défis pour l'organisation des cours en présentiel, notamment des problèmes logistiques (dé- placements entre les campus de Boulogne-Billancourt et de Saint-Germain-des-Prés, et manque de salles disponibles à Boulogne-Billancourt). Les cours magistraux (CM) mutualisés, dispensés dans des amphithéâtres de 200 à 300 étudiants, ont également montré leurs limites :

- Faiblesse des interactions entre les étudiants et l'enseignant ;
- Apprentissage passif des étudiants, car la transmission de la connaissance est unidirectionnelle et descendante de l'enseignant vers les étudiants ;
- Manque de personnalisation, car le format magistral ne permet pas de répondre de manière personnalisée aux besoins ou aux questions spécifiques des étudiants.

En outre, de nombreux étudiants exercent une activité rémunérée au cours de leur licence (54 % des apprenants selon Henri-Panabière, Lejeune et Vacant, 2023), ce qui leur permet de se familiariser avec les métiers de l'éducation tout en finançant leurs études via des dispositifs comme l'Emploi Avenir Professeur (AEP) et la « préprofessionnalisation ». Ces contraintes ont motivé une révision des modalités de formation pour alléger les contraintes temporelles et logistiques. Ainsi, l'introduction des outils numériques, notamment via une plateforme LMS (Learning Management System), a été envisagée pour pallier ces problèmes. Cependant, une formation entièrement à dis- tance présente des risques, tels que l'affaiblissement du cadre pédagogique et l'augmentation du travail non encadré pour les étudiants. Depuis 2018-2019, les formations en sciences de l'éducation sont donc hybrides, avec une distinction entre les cours adaptés à cette modalité et ceux nécessitant une présence physique. Les cours magistraux (CM) sont principalement dispensés à distance, tandis que les travaux dirigés (TD) sont maintenus en présentiel. La proportion des cours en présentiel diminue progressivement de la première à la troisième année de licence : L1 = 72 % des heures de cours en présentiel, L2 = 58 % de cours en présentiel, L3 = 51 % de cours en présentiel. Il con- vient aussi de préciser que les étudiants consacrent en moyenne 7 h 25 min de volume horaire par semaine aux activités d'apprentissage hors face-à-face. En 2023-2024, la formation hybride en sciences de l'éducation en est à sa sixième année. Après six ans d'existence de ce dispositif, il est pertinent d'examiner les apprentissages des étudiants dans cet environnement hybride. Il convient de préciser que les étudiants participant à la présente étude ont intégré la formation avec le dispositif hybride depuis un (licence 1), deux (licence 2) ou trois (licence 3) ans.

Pour réussir leurs études, les étudiants mettent en œuvre des apprentissages autorégulés. Selon Allal (2011), l'autorégulation est une qualité intrinsèque de chaque apprenant. Les recherches montrent que ceux qui réussissent en ligne manifestent cette compé- tence (Cosnefroy, 2019, 2022 ; Depover et al., 2016 ; Djiezion et Bernard, 2022, 2023 ; Djiezion, 2021 ; Sauvé et al., 2007).

Ainsi, en général, les apprentissages autorégulés contribuent au succès scolaire. Les étudiants qui développent ces compétences réussissent mieux. Cependant, qu'en est-il de l'évolution de cette variable dans un dispositif hybride ? Cette étude explore l'évolution des apprentissages autorégulés en fonction de l'expérience dans un contexte de formation hybride universitaire

3.2. Hypothèses de recherche

Dans leur recherche sur les environnements d'apprentissage distanciels et hybrides, Sahin et Shelley (2008) ont mis en exergue les effets de l'expérience sur les apprentissages autorégulés. Leur recherche a montré que plus les étudiants s'engagent dans des environnements mixtes ou distanciels, plus ils développent des stratégies d'autorégulation qui renforcent leur sentiment d'efficacité personnelle. Broadbent (2017) a observé que les étudiants débutants dans les dispositifs hybrides ou en ligne ont souvent du mal à s'adapter au besoin de réguler leur apprentissage. Toutefois, avec l'expérience, ils développent des stratégies plus efficaces de gestion des ressources, d'organisation des tâches et d'ajustement de leur niveau de concentration. Sun et Rueda (2012) ont constaté que les étudiants plus expérimentés dans les cours en ligne démontrent de meilleures compétences d'autosurveillance et d'autoévaluation, ce qui leur permet de s'adapter aux exigences d'une formation autonome.

Sur la base de ces recherches, nous posons l'hypothèse suivante : plus l'apprenant a de l'expérience dans un contexte de formation hybride universitaire, plus son niveau d'apprentissages autorégulés est élevé.

4 Méthodologie et résultats de l'étude

Pour cette recherche quantitative, nous avons choisi d'utiliser une approche méthodologique basée sur une démarche hypothético-déductive. Cette méthode commence par la formulation d'un problème ou d'une question de recherche, suivie par une ou plusieurs hypothèses à tester et à valider, ou à rejeter, par le biais d'une procédure scientifique expérimentale (Livian, 2015). Cette section se concentrera donc sur la présentation de la méthodologie employée dans le cadre de notre étude quantitative.

4.1 Population cible

Nous souhaitons en principe appliquer les résultats à la population globale visée par l'étude. Pour cette recherche, cela inclut les étudiants inscrits en licence de sciences de l'éducation et de la formation de l'Université Paris Cité durant l'année universitaire 2023-2024. Ainsi, 268 étudiants sont inscrits en licence de sciences de l'éducation de l'Université Paris Cité.

Nous avons sélectionné deux critères pour définir cette population : le niveau scolaire et le parcours scolaire. Nous avons trois niveaux scolaires (licences 1, 2 et 3) et deux parcours (professorat des écoles et lien social). Nous estimons que ces critères permettent de mieux décrire la population cible de l'étude. Selon ces critères, cette population se répartit comme suit (tableau 1).

Tableau 1. Répartition de la population cible

Catégorie	Nombre d'individus	Pourcentage	Hommes (9 %)	Femmes (91 %)
Total de la population	268	100%	24	244
L1	83	31%	7	76
L2	93	34,70%	8	85
L3	92	34,30%	8	84
Professorat des écoles	172	64,20%	15	157
Lien social	96	35,80%	9	87

4.2 Échantillonnage

Un échantillon est un sous-ensemble plus petit de la population d'origine, sélectionné par des méthodes probabilistes ou non, afin de refléter au mieux cette population (Savard, 1978). Ainsi, l'étude est réalisée sur une partie de cette population initiale. En supposant que l'échantillon est représentatif, les résultats de l'étude peuvent être généralisés à l'ensemble de la population grâce à la statistique inférentielle. En effet, la statistique inférentielle permet de projeter les résultats obtenus à partir de l'échantillon sur toute la population concernée avec un certain degré de certitude (Spiegel, 1974).

Pour pouvoir généraliser les résultats à la population d'origine, il est crucial que l'échantillon soit représentatif, et la méthode d'échantillonnage joue un rôle clé dans cette représentativité. Il existe des méthodes d'échantillonnage probabilistes et non probabilistes. Dans le cadre d'un échantillonnage probabiliste, chaque membre de la population d'origine a une probabilité égale d'être inclus dans l'échantillon. Pour appliquer cette méthode, il est indispensable de disposer d'une base de données répertoriant tous les individus de la population concernée. Dans cette recherche, cette condition est remplie. En effet, les étudiants constituant la population source sont connus, puisque le département de sciences de l'éducation de la faculté des sciences humaines de l'Université Paris Cité possède une base de données complète de ses étudiants. C'est donc à partir de cette base de données que le questionnaire a été envoyé en ligne.

Nous faisons remarquer que dans le cadre de la présente étude, il s'agit d'un échantillonnage probabiliste. En effet, L'échantillonnage probabiliste repose sur l'idée que chaque individu de la population a une probabilité connue et non nulle d'être sélectionné. Le biais de sélection est ainsi réduit, ce qui augmente la généralisabilité des résultats, même si tous les individus ne répondent pas.

Le questionnaire a été rempli en ligne grâce à un lien envoyé à chaque étudiant par courriel durant le mois d'avril 2024. Le choix des répondants a ainsi été basé sur les réponses reçues via la plateforme de sondage en ligne. Dans ce contexte, il n'était pas nécessaire de procéder à un tirage au sort préalable, car le questionnaire a été envoyé à l'ensemble de la population d'origine, et ceux qui ont répondu de manière ont constitué l'échantillon. Ainsi, 268 questionnaires ont été envoyés en ligne.

A la suite de cette démarche, 110 réponses ont été recueillies. Ce nombre correspond à un taux de réponse de 41 %. Le taux de réponse atteint (41 %) s'inscrit dans les standards méthodologiques reconnus pour les enquêtes en sciences humaines et sociales, où les taux de participation oscillent habituellement entre 20 % et 50 % (Observatoire Espace et Société, s.d.; SurveyMonkey, s.d.). Ce niveau de réponse contribue à garantir la robustesse et la représentativité des données collectées.

En tenant compte des critères de répartition de la population mère que sont le niveau d'études et le parcours professionnel, nous obtenons le tableau suivant sur le plan de l'échantillon (tableau 2).

Tableau 2. Répartition de l'échantillon

Catégorie	Nombre d'individus	Pourcentage	Hommes (19,2 %)	Femmes (80,8 %)
Total de la population	110	100%	21	89
L1	39	35,6 %	7	32
L2	29	26 %	6	23
L3	42	38,4 %	8	34
Professorat des écoles	75	68,5 %	14	61
Lien social	35	31,5 %	7	28

A la vue du tableau 3 qui compare les répartitions population mère et échantillon, nous estimons que la répartition des répondants est assez proche de celle des inscrits selon les critères de niveau et de parcours. On observe donc une certaine représentativité structurelle de l'échantillon par rapport à la population de référence.

Tableau 3. Comparaison des répartitions population mère et échantillon

Catégorie	Participants (268)	Répondants (110)
Total de la population	268	110
Hommes	24 (9 %)	21 (19,2 %)
Femmes	244 (91 %)	89 (80,8 %)
L1	83 (31 %)	39 (35,6 %)
L2	93 (34,7 %)	29 (26 %)
L3	92 (34,3 %)	42 (38,4 %)
Professorat des écoles	172 (64,2 %)	75 (68,5 %)
Lien social	96 (35,8 %)	35 (31,5 %)

4.3 Traitement statistique

Après avoir recueilli les réponses à l'aide de la plateforme de sondage en ligne Sphinx Déclic, les données ont été téléchargées au format CSV et importées dans le logiciel Stata 12.1 pour analyse. Tout d'abord, des tableaux thématiques ont été créés en regroupant les réponses aux différentes questions selon des tris à plat. Ensuite, nous avons utilisé l'échelle des apprentissages autorégulés en ligne (EAREL) (Cosnefroy, Fenouillet et Heutte, 2020) afin d'obtenir les indices d'apprentissages autorégulés en ligne (IAREL) (Cosnefroy, Fenouillet et Heutte, 2020) de chacun des 110 participants à l'étude. Ce sont ces deux échelles qui ont fait l'objet du calcul de coefficient de corrélation.

4.4 L'échelle de l'autorégulation des apprentissages dans des environnements en ligne (EAREL)

Afin d'évaluer le niveau d'apprentissages autorégulés des étudiants en licence de sciences de l'éducation de l'Université Paris Cité, nous nous sommes servis d'une échelle dénommée : échelle de l'autorégulation des apprentissages dans des environnements en ligne (EAREL). Pour les créateurs Cosnefroy, Fenouillet et Heutte (2020), il s'agit d'un outil permettant d'évaluer le niveau de l'engagement des étudiants dans le cadre de formations en ligne. L'échelle comprend 24 items, répartis en 4 sous-échelles de 6 items chacune : stratégies d'apprentissage (COGN), contrôle du contexte d'apprentissage (CTXTE), recherche de soutien entre pairs (PAIRS) et procrastination (PROC(r)) (Cosnefroy, Fenouillet et Heutte, 2018).

Les 24 énoncés de l'EAREL sont répartis en 6 items par sous-échelle et sont évalués sur une échelle graduée de 1 à 7. L'analyse des résultats de l'étude à partir de l'EAREL a permis de calculer les indices d'apprentissage autorégulé en ligne (IAREL) de chaque participant. Il s'agit des indices suivants : indice de procrastination, indice de soutien des pairs, indice de stratégie d'apprentissage et indice de contrôle du contexte d'apprentissage (Cosnefroy, Fenouillet et Heutte, 2020). Chaque indice est constitué de 6 items repartis de la manière suivante (tableau 4).

Tableau 4. Indices et items correspondants (confère annexe)

Item	Numéros des items correspondants
Soutien des pairs (PAIRS)	4+5+10+13+17+24
Stratégie d'apprentissage (COGN)	1+9+14+18+21+22
Contrôle du contexte d'apprentissage (CTXTE)	3+6+11+16+20+8
Procrastination (PROC)	2+12+15+19+7+23

Les indices ont été obtenus de la manière suivante : nous avons fait les moyennes des scores des items correspondant à chaque indice. Pour avoir ces indices, nous avons calculé le score de chaque modalité à la suite des réponses données par les 110

répondants. Ces répondants devaient en effet donner une appréciation à chaque modalité sur une échelle allant de 1 (pas du tout) à 7 (tout à fait).

4.5 Coefficient de corrélation entre les apprentissages autorégulés et l'expérience du contexte de formation hybride.

Le coefficient de corrélation, souvent noté r, mesure la force et la direction de la relation linéaire entre deux variables. Il détermine l'intensité de la corrélation entre deux variables x et y.

Le coefficient de corrélation de Pearson est utilisé pour les données bivariées continues afin de déterminer l'intensité et le sens de la corrélation linéaire entre les deux ensembles de données. Voici une interprétation détaillée de ce coefficient :

Valeurs du coefficient de corrélation

- r = 1 : Corrélation parfaitement positive. Les deux variables augmentent ou diminuent ensemble de manière linéaire.
- r = −1 : Corrélation parfaitement négative. Une variable augmente quand l'autre dimi- nue, et cette relation est parfaitement linéaire.
- r = 0 : Aucune corrélation linéaire. Il n'y a pas de relation linéaire entre les deux variables.

Interprétation des valeurs de r :

- 0,9 à 1 ou -0,9 à -1 : Corrélation très forte.
- 0,7 à 0,9 ou -0,7 à -0,9 : Corrélation forte.
- 0,5 à 0,7 ou -0,5 à -0,7 : Corrélation modérée.
- 0,3 à 0,5 ou -0,3 à -0,5 : Corrélation faible.
- 0 à 0,3 ou 0 à -0,3 : Corrélation très faible.

Dans le cadre de la présente recherche, les variables dont nous allons analyser la corrélation avec l'expérience en contexte de formation hybride sont Les scores de l'échelle des apprentissages autorégulés en ligne (EARL).

5 Résultats de l'étude et vérification des hypothèses

5.1 Explications

Nous avons divisé les apprenants en deux groupes distincts. Nous avons, d'une part, les étudiants qui découvrent le dispositif de formation hybride universitaire, notamment les étudiants en licence 1 de sciences de l'éducation de l'Université Paris Cité et, d'autre part, les étudiants qui ont une expérience du dispositif de formation hybride universitaire, notamment les étudiants en licence 2 et 3 de sciences de l'éducation de l'Université Paris Cité. Nous avons donc calculé la moyenne des indices d'apprentissage auto- régulés de chaque groupe et nous avons établi le pourcentage d'évolution. Puis nous avons fait un test de Student afin de déterminer si les différences sont statistiquement pertinentes. En effet, le test de Student à deux échantillons est une méthode statistique utilisée pour comparer les moyennes de deux groupes

indépendants. Ce test permet de déterminer si les différences observées entre les moyennes de ces deux échantillons sont statistiquement significatives ou si elles pourraient être dues au hasard.

Son interprétation est la suivante :

- Si la p-valeur est inférieur à un seuil de 0,05, on rejette l'hypothèse nulle. Cela signifie qu'il existe une différence statistiquement significative entre les deux groupes ;
- Si la p-valeur est supérieure au seuil choisi, on ne peut pas rejeter l'hypothèse nulle, ce qui signifie que la différence observée entre les groupes n'est pas significative et pour- rait être due au hasard.

5.2 Expériences du dispositif de formation hybride universitaire et apprentissages autorégulés

Analysons maintenant les résultats du test de Student à deux échantillons pour comparer les indices d'apprentissages, tels que présentés au tableau 5.

Tableau 5. Moyennes comparées des indices d'apprentissages autorégulés en ligne (IAREL) des étudiants

Indice des apprentissages autorégulés en ligne	Moyenne des indices : étudiants débutants (licence 1)	Moyenne des indices : étudiants expérimentés (licences 2 et 3)	p-value
Soutien des pairs (PAIRS)	3,76	3,98	0,02 (< 0,05)
Stratégie d'apprentissage (COGN)	4,39	4,62	0,04 (< 0,05)
Contrôle du contexte d'apprentissage (CTXTE)	5,18	5,57	0,001 (< 0,05)
Procrastination (PROC)	4,04	4,3	0,45 (> 0,05)

Soutien des pairs (PAIRS)

p-value = 0,02 (< 0,05)

La différence observée (3,76 vs 3,98) est statistiquement significative. Cela suggère que les étudiants expérimentés bénéficient d'un soutien des pairs plus important que les débutants.

Stratégie d'apprentissage (COGN)

p-value = 0,04 (< 0,05)

La différence observée (4,39 vs 4,62) est significative. Cela indique que les stratégies d'apprentissage sont mieux développées chez les étudiants expérimentés.

Contrôle du contexte d'apprentissage (CTXTE)

p-value = 0,001 (< 0,05)

La différence est très significative (5,18 vs 5,57). Cela montre que les étudiants expérimentés ont un meilleur contrôle de leur environnement d'apprentissage.

Procrastination (PROC)

p-value = 0,45 (> 0,05)

La différence observée (4,04 vs 4,3) n'est pas statistiquement significative. Cela signifie que, bien que les étudiants expérimentés semblent légèrement moins procrastiner, cette différence n'est pas assez marquée pour être considérée comme significative. La procrastination semble être un défi relativement constant pour les deux groupes.

Ces résultats confirment l'hypothèse suivante : plus l'apprenant a de l'expérience dans un contexte de formation hybride universitaire, plus son niveau d'apprentissages auto-régulés est élevé.

6 Discussion, limites et perspectives de l'étude

Les résultats de cette étude soulignent l'impact positif de l'expérience des étudiants dans les dispositifs de formation hybride sur le développement de leurs apprentissages autorégulés. Ces conclusions s'appuient sur l'analyse comparative des indices d'apprentissages autorégulés en ligne (IAREL) entre les étudiants débutants (licence 1) et les étudiants expérimentés (licences 2 et 3).

Les indices d'apprentissages autorégulés en ligne révèlent des différences significatives entre les étudiants débutants et expérimentés, corroborant l'hypothèse de départ selon laquelle plus les étudiants sont expérimentés, plus leurs compétences d'apprentissages autorégulés se développent. Les différences significatives observées dans les indices d'apprentissages autorégulés en ligne entre les étudiants débutants et expérimentés (soutien des pairs, stratégie d'apprentissage, contrôle du contexte d'apprentissage) corroborent les travaux d'Artino (2008), qui a démontré que les étudiants plus expérimentés dans des environnements d'apprentissage en ligne développent des compétences d'autorégulation plus solides. Artino a souligné l'importance de l'expérience dans l'ajustement des stratégies d'apprentissage et dans la gestion efficace de l'environnement d'apprentissage. Zimmerman (2002) a mis en avant l'idée que l'autorégulation est un processus évolutif qui s'améliore avec l'expérience. Nos résultats sur l'amélioration des stratégies d'apprentissage et du contrôle du contexte d'apprentissage s'alignent sur cette perspective.

On voit le rôle de l'expérience, mais d'une expérience couronnée de succès (les étudiants ont réussi à accéder en L3). L'expérience joue un rôle central dans les AAR, mais toutes les expériences ne contribuent pas de manière égale à son renforcement. Comme le soulignent de nombreux travaux en psychologie sociale et éducative (Pajares, 2006, Zimmerman et Schunk, 2011), ce sont les expériences couronnées de succès qui ont le plus grand impact sur la construction des apprentissages autorégulés. Dans le cas étudié, les étudiants qui ont réussi à accéder en L3 illustrent bien cet effet.

Leur parcours témoigne d'une capacité à surmonter des obstacles et à atteindre des objectifs scolaires, renforçant ainsi leur capacité d'autorégulation dans un contexte universitaire hybride.

Il faut aussi noter que les étudiants en L1 se trouvent dans une phase charnière marquée par une transition majeure. Pour les jeunes adultes, il s'agit généralement du passage du lycée à l'université, un moment où les repères éducatifs, sociaux et parfois personnes sont bouleversés. Pour les adultes plus âgés qui retournent aux études, cette transition prend une autre forme, souvent marquée par un réajustement identitaire face aux exigences scolaires et parfois à des responsabilités familiales ou professionnelles.

Ces transitions, bien qu'enrichissantes, sont également génératrices d'incertitudes. Les étudiants doivent faire face à des attentes nouvelles, à des méthodes d'enseignement différentes, et souvent à une autonomie accrue dans leur apprentissage. Cette incertitude peut entraîner une fluctuation des AAR, car les apprenants n'ont pas encore eu l'opportunité de valider leur capacité à s'adapter et à réussir dans ce nouvel environnement.

7 Conclusion

Cette étude visait à explorer les expériences des étudiants dans un dispositif de formation hybride universitaire et à examiner leur impact sur les apprentissages autorégulés. Les résultats montrent une évolution positive chez les étudiants expérimentés (licences 2 et 3) par rapport aux étudiants débutants (licence 1) en ce qui concerne plusieurs dimensions des apprentissages autorégulés, telles que le soutien des pairs, les stratégies d'apprentissage et le contrôle du contexte d'apprentissage.

Cette recherche a toutefois quelques limites qu'il convient de souligner. L'étude se base sur des comparaisons transversales entre deux groupes d'étudiants (débutants et expérimentés), ce qui ne permet pas d'observer l'évolution des individus au cours du temps. Une approche longitudinale offrirait une meilleure vue d'ensemble sur la progression de leurs apprentissages autorégulés et de leur sentiment d'efficacité personnelle. En outre, d'autres facteurs, tels que le soutien institutionnel, la motivation intrinsèque des étudiants ou encore la qualité des dispositifs de formation, peuvent influencer les résultats. Ces aspects n'ont pas été spécifiquement pris en compte dans cette étude et pouraient avoir un impact sur la régulation des apprentissages. Enfin, cette étude n'établit pas, par exemple, de relations entre le contenu et le format des UE et l'évolution des apprentissages autorégulés de étudiants dans le temps.

Les limites de l'étude ouvrent de nouvelles perspectives pour approfondir l'étude, envisager des prolongements ou améliorer d'autres recherches sur des thématiques similaires. Une première piste consisterait à adopter une approche longitudinale afin de suivre l'évolution des apprentissages autorégulés et du sentiment d'efficacité personnelle sur une période prolongée. Cela permettrait de dépasser les limites des comparaisons transversales et d'observer les dynamiques individuelles dans le temps. De plus, intégrer des variables contextuelles telles que le soutien institutionnel, la motivation intrinsèque ou encore la qualité perçue des dispositifs de formation offrirait une compréhension plus complète des facteurs influençant la régulation des apprentissages.

Par ailleurs, il serait pertinent d'explorer les liens entre les caractéristiques des unités d'enseignement — notamment leur contenu, leur format (présentiel, hybride, distanciel) et leur mode d'évaluation — et le développement des stratégies d'apprentissage autorégulé. Une telle approche permettrait d'identifier les conditions pédagogiques les plus favorables à l'autonomisation des étudiants, et de nourrir la conception de dispositifs plus adaptés au développement de ces compétences clés dans l'enseignement supérieur.

Références

1. Ali, A., Khan, R. M. I., & Alouraini, A. (2023). A Comparative Study on the Impact of Online and Blended Learning. SAGE Open, 13(2), 21582440231154417.
2. https://doi.org/10.1177/21582440231154417
3. Allal, L. (2011). Pedagogy, didactics and the co-regulation of learning: a perspective from the French-language world of educational research. Research Papers in Education, 26(3), 329-336. http://doi.org/10.1080/02671522.2011.595542
4. Artino, A. R. (2008). Motivational beliefs and perceptions of instructional quality: Predicting satisfaction with online training.
5. Journal of Computer Assisted Learning, 24(3), 260-270. https://doi.org/10.1111/j.1365- 2729.2007.00258.x
6. Broadbent, J. (2017). Comparing online and blended learner's self-regulated learning strategies and academic performance. The Internet and Higher Education, 33, 24-32.
7. https://doi.org/10.1016/j.iheduc.2017.01.004
8. Charlier, B., Deschryver, N. et Peraya, D. (2006). Apprendre en présence et à distance : une définition des dispositifs hybrides. Distances et savoirs, 4, 469-496. https://shs.cairn.info/revue- distances-et-savoirs-2006-4-page-469
9. Cosnefroy, L. (2010). L'apprentissage autorégulé : perspectives en formation d'adultes. Savoirs, 23, 9-50. https://doi.org/10.3917/savo.023.0009
10. Cosnefroy, L. (2019). Chapitre 9. L'autorégulation des apprentissages et la e-Formation. Dans
11. Jézégou (dir.), Traité de la e-Formation des adultes (p. 213-232). De Boeck. https://doi.org/10.3917/dbu.jezeg.2019.01.0213
12. Cosnefroy, L. (2022, 6 janvier). Les spécificités de l'apprentissage autorégulé en e-Formation [webinaire]. Lille.Pod. https://pod.univ-lille.fr/
13. Cosnefroy, L., Fenouillet, F et Heutte, J. (2018, 21-23 mars). Développement et validation d'une échelle d'apprentissage autorégulé en ligne. 2e Colloque international e-Formation des Adultes et Jeunes Adultes, Lille, France.
14. Cosnefroy, L., Fenouillet, F., et Heutte, J. (2020). Construction et validation de l'Échelle d'auto- régulation des apprentissages en ligne (EAREL). Canadian Journal of Behavioural Science / Re- vue canadienne des sciences du comportement, 52(3), 255–260. https://psycnet.apa.org/doi/10.1037/cbs0000147
15. Depover, C., Mélot, L., Strebelle, A. et Temperman, G. (2016). Régulation et autorégulation dans les dispositifs d'apprentissage à distance. Dans B. Noel et C. Cartier (dir.), De la métacognition à l'apprentissage autorégulé (p. 95-110). De Boeck.
16. Djiezion, E. G. (2021). Interactions hétéro et homo régulées et apprentissages autorégulés en dispositif universitaire de formation et de communication médiatisées [thèse de doctorat, Uni- versité Paris Cité, France]. HAL theses. https://theses.hal.science/tel-03573035

17. Djiezion, E. G. et Bernard, F.-X. (2022). Régulations externes et apprentissages autorégulés en université virtuelle.
18. Médiations et médiatisations, (10), 26-44. https://doi.org/10.52358/mm.vi10.334
19. Djiezion, E. G. et Bernard, F.-X. (2023). Corrélation entre la fréquence des interactions et le niveau d'autorégulation des étudiants et étudiantes en dispositif universitaire de formation en ligne. Revue internationale des technologies en pédagogie universitaire, 20(3), 73-89.
20. https://doi.org/10.18162/ritpu 2023-v20n3-05
21. Garrison, D. R. et Kanuka, H. (2004). Blended learning: Uncovering its transformative potential in higher education. The Internet and Higher Education, 7(2), 95-105. https://doi.org/10.1016/j.iheduc.2004.02.001
22. Henri-Panabière, G., Lejeune, G. et Vacant, J. (2023, décembre). Retour sur l'hybridation des enseignements en Licence de Sciences de l'éducation : expériences et organisation de la vie étu- diante. Université Paris Cité.
23. Livian, Y. (2015, 12 janvier). Initiation à la méthodologie de la recherche en SHS : réussir son mémoire ou thèse. Archives ouvertes (HAL). https://shs.hal.science/halshs-01102083/document.
24. Means, B., Toyama, Y., Murphy, R., Baki, M. et Jones, A. (2009). Evaluation of evidence-based practices in online learning: A meta-analysis and review of online learning studies. U.S. Department of Education.
25. Pajares, F. (2006). Self-efficacy during childhood and adolescence: Implications for teachers and parents. Dans F. Pajares et T. Urdan (dir.), Self-efficacy beliefs of adolescents (pp. 339-367). Information Age Publishing.
26. Paquelin D. et Choplin, H., (2003). Du prescrit au vécu, l'enjeu des régulations. Dans B. Albero (dir.), Autoformation et enseignement supérieur. Hermes Science.
27. Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning. Dans M. Boekaerts,
28. P. R. Pintrich et
29. M. Zeidner (dir.), Handbook of self-regulation (p. 451-502). Academic Press.
30. Rayyan, M., Abusalim, N., Alshanmy, S., Abu Awwad, F. M., & Alghazo, S. (2024). Enhancing Academic Performance through Blended Learning: A Study on the Relationship between Self- Efficacy and Student Success. International Journal of Interactive Mobile Technologies (iJIM), 18(19), 52-67. https://doi.org/10.3991/ijim.v18i19.49835
31. Sahin, I. et Shelley, M. (2008). Considering students' perceptions: The distance education student satisfaction model. Educational Technology & Society, 11(3), 216-223.
32. Sauvé, L., Debeurme, G., Martel, V., Wright, A., Hanca, G. et Castonguay, M. (2007, juin). SAMI-Persévérance. L'abandon et la persévérance aux études postsecondaires - Rapport final. Québec : Rapport déposé au FQRSC. http://www.savie.qc.ca/CampusVirtuel/Upload/Fi- chiers/SAMI-_rapport_final_vf.pdf
33. Savard, J. G. (1978) Statistiques (N. Gilbert, trad. et adaptation).
34. Schwarzer, R. et Jerusalem, M. (1995). Generalized Self-Efficacy scale. Dans J. Weinman, S. Wright et M. Johnston, Measures in health psychology: A user's portfolio. Causal and control beliefs (p. 35-37). Windsor, England: NFER- Nelson.
35. Spiegel, M.R. (1974). Theorie et applications de la statistique, Ediscience.
36. Sun, J. C.-Y. et Rueda, R. (2012). Situational interest, computer self-efficacy and self-regulation: Their impact on
37. student engagement in distance education. British Journal of Educational Technology, 43(2), 191-204. https://doi.org/10.1111/j.1467-8535.2010.01157.x

38. Zimmerman, B. J. et Schunk, D. H. (2011). Handbook of self-regulation of learning and performance. Routledge.
39. Zimmerman, B. J. (2002). Becoming a Self-Regulated Learner: An Overview, Theory Into Practice, 41(2), 64-70.

Des compétences aux savoir-faire : proposition d'extensions au métamodèle de COMPER et à MOODLE

Patrick Etcheverry[1 2], Thierry Nodenot[1 2], Marta Toribio[2], Christophe Marquesuzaà[1 2], Jennifer Lesbats[2], Jose Manuel Cornejo[1 2], Benjamin Fontaine[1]

[1] Université de Pau et des Pays de l'Adour, laboratoire LIUPPA, Anglet, France
prenom.nom@iutbayonne.univ-pau.fr
[2] Université de Pau et des Pays de l'Adour, IUT de Bayonne, Anglet, France
prenom.nom@iutbayonne.univ-pau.fr

Résumé. Nous proposons dans cet article un modèle conceptuel et une solution technique visant à offrir des services support à l'approche par compétences (APC) tirant parti des résultats du projet COMPER. Considérant les apports de l'APC attendus tant pour les étudiants (autonomie / auto-régulation) que pour les enseignants, nous proposons des extensions aux métamodèles de COMPER et de Moodle centrées sur le concept de savoir-faire (« *skill* »). Nous présentons une opérationnalisation de ce modèle dans Moodle ainsi que les intérêts pour l'APC d'une telle intégration.

Mots-clés : APC, COMPER, LMS Moodle, Modélisation des connaissances et des savoir-faire, Objets de recommandation, Graphe de compétences

Abstract. We propose in this article a conceptual model and a technical solution aimed at providing support services for the competency-based approach (CBA) leveraging the results of the COMPER project. Considering the effects of the CBA aimed for both students (autonomy/self-regulation) and teachers, we propose extensions to the COMPER and Moodle metamodels centered on the concept of 'skill.' We present an operationalization of this model in Moodle as well as the educational benefits for CBA of such an integration.

Keywords: CBA, COMPER, LMS Moodle, Representation of knowledge and skills, Recommendation subjects, Skills graph

1 Introduction

La notion de Compétence a toujours fait l'objet d'une attention particulière de la communauté scientifique internationale en Sciences et Technologies de l'Information pour l'Education et la Formation [1][2][3][4][5]. En France, l'introduction en 2019 de l'APC pour le suivi et l'évaluation des étudiants dans l'Enseignement Supérieur a motivé un nouveau flux de travaux [6] à dimensions théorique et pratique ; le nouveau

Programme National des 24 spécialités du récent Bachelor Universitaire de Technologie (BUT) porte ainsi un défi important pour les équipes enseignantes ; les compétences y occupent un rôle central et chaque unité d'enseignement participe à l'acquisition d'une ou plusieurs compétences définies dans le programme.

Pour construire ce programme, les I.U.T. ont été accompagnés par le Laboratoire de soutien aux Synergies Éducation Technologies (LabSET) [7] de l'Université de Liège. Ce laboratoire, reconnu pour ses travaux en ingénierie pédagogique, s'est appuyé sur le cadre théorique de l'approche par compétences (APC) défini par Jacques Tardif [8] qui définit une compétence comme « *[un] savoir-agir complexe reposant sur la mobilisation et la combinaison efficaces d'une variété de ressources internes et externes à l'intérieur d'une famille de situations* ». L'idée de « *savoir-agir*» donne un caractère englobant à la compétence et évite que cette dernière soit réduite à un ensemble de savoir-faire ou de connaissances procédurales.

Afin de pouvoir mobiliser ces savoirs agir complexes via Moodle, nous (ingénieurs pédagogiques et chercheurs) avons exploré les fonctionnalités d'exploitation des compétences intégrées dans ce LMS. Nous avons donc manipulé des référentiels de compétences (hiérarchisées [9]) à l'échelle d'un ensemble de cours et de ressources Moodle, ce qui nous a amené à approfondir les fonctionnalités de Moodle sur ce sujet spécifique ; les fonctionnalités offertes donnent des possibilités de validation et de suivi destinées aux enseignants ainsi qu'aux étudiants. Cela permet notamment à l'étudiant d'avoir un regard sur les compétences travaillées, validées et celles qui restent à développer. Ces informations lui permettent de se situer dans son apprentissage et d'enclencher un processus d'autorégulation [10], c'est-à-dire, choisir de manière autonome sa prochaine action d'apprentissage après autoévaluation de ses échecs / réussites. La gestion des compétences n'est donc pas une brique en plus qui serait indépendante de toutes les autres fonctionnalités de Moodle, mais dorénavant tout cours, toute ressource et toute activité Moodle inclut une facette « Compétences » que l'enseignant et l'étudiant peuvent exploiter lors de l'édition de contenus, mais aussi lors du suivi et de l'évaluation des étudiants.

Selon nous, il n'y a donc pas lieu de vouloir faire table rase de ces apports encore assez peu exploités par la communauté éducative. Et ce même si nous constatons avec [11] qu'« *... au bout d'un moment, ..., on n'est plus au niveau d'une compétence, mais que l'on est au niveau des savoirs et savoir-faire* » (page 72 de l'article référencé). Au contraire, pour mettre en œuvre l'APC dans leurs pratiques, enseignants et étudiants doivent pouvoir exploiter d'autres éléments inexistants à ce jour dans le métamodèle des LMS (et en particulier de Moodle), car le niveau de granularité de description de leurs séquences d'enseignement ne peut se suffire du concept de Compétence au sens de Moodle : savoirs et savoir-faire doivent aussi pouvoir être précisément décrits et exploités tant par les étudiants que les enseignants [12][13].

Dans ce contexte, la question centrale qui guide notre travail est la suivante : comment ancrer l'APC dans la réflexion des enseignants lors de l'élaboration et de l'édition des activités et ressources constituant un cours en ligne ?

Partant d'un cas d'étude, nous présentons dans la section suivante les éléments d'un modèle permettant de décrire des compétences de niveau micro (cf. savoirs et savoir-faire) en croisant les concepts de COMPER [11] [14] avec ceux de Moodle. Nous proposons dans la section 3 une opérationnalisation de ce modèle via une extension Moodle. Enfin, nous concluons par une présentation de nos travaux en cours visant à s'appuyer sur cette approche par compétences pour promouvoir des services d'accompagnement des étudiants.

2 De la nécessité d'étendre le métamodèle de COMPER

2.1 Expression des besoins de modélisation

Dans le référentiel du BUT Informatique, une compétence clé à acquérir en première année est « Développer des applications informatiques simples ». Cette compétence globale décrit un objectif d'apprentissage important, mais elle ne détaille pas toutes les concepts seuils et savoir-faire nécessaires pour atteindre ce niveau de maîtrise. Par exemple, l'exercice « simple » visant à élaborer un programme/application calculant le périmètre d'un cercle sous-entend la compréhension et la maîtrise de plusieurs concepts et savoir-faire (que nous désignons par le terme « *skill* » dans la suite de cet article). En termes de concepts, les étudiants doivent par exemple savoir ce qu'est une variable, le type d'une variable, la valeur d'une variable ou encore ce qu'est une affectation. En termes de savoir-faire, les étudiants devront être capables de nommer une variable, typer une variable, initialiser une variable, déclarer une variable, etc.

Pour accompagner les étudiants sur cet exercice, l'enseignant devra peut-être ramener les étudiants sur des extraits de cours présentant des définitions, des exemples ou des tutoriels en lien avec ces concepts (qui sont de fait des savoirs) et ces skills. Bien que Moodle permette de modéliser des compétences globales telles que « Développer des applications informatiques simples » et de les hiérarchiser, il n'offre pas d'outils pour modéliser correctement ces concepts et ces skills et éventuellement pointer des ressources de cours associées pour aider les étudiants.

Sur le sujet spécifique de la représentation des compétences, des savoirs et des savoir-faire, les propositions relevant de COMPER constituent un corpus de travaux récents présentant un grand intérêt. COMPER (COMpétences et PERsonnalisation) est à la fois le nom d'un projet ANR et celui d'un métamodèle [11] visant à produire des outils et des services appuyés sur l'APC dans une perspective de personnalisation et d'(auto)régulation des apprentissages, et ce quelle que soit la plateforme numérique utilisée. Les outils développés qui ont fait l'objet de campagnes d'expérimentation permettent tout à la fois de définir un référentiel de compétences, de savoirs et savoir-faire, d'associer des ressources pédagogiques à des éléments de ce référentiel, d'établir / d'éditer un profil de compétences des apprenants [15] et de faciliter la régulation par des enseignants mais surtout l'autorégulation par les apprenants eux-mêmes [16].

Pour ce qui est de la modélisation du référentiel de compétences, le métamodèle fournit un vocabulaire normalisé et des éléments de structuration des compétences, des savoirs et des savoir-faire. On retrouve dans le métamodèle la capacité d'exprimer

qu'un savoir ou savoir-faire est issu d'une compétence, de décrire finement des savoirs et savoirs faire via des relations diversifiées, de relier savoirs et savoir-faire à des ressources pédagogiques, chacune de ces dernières étant caractérisée par une simple url afin de garder une indépendance vis-à-vis des LMS.

En nous appuyant sur les concepts et les relations du métamodèle COMPER, nous avons essayé de modéliser notre cas d'étude et de l'implanter dans Moodle, ce qui nous a amené aux constats suivants :

- Alors que Moodle permet de relier toute compétence d'un référentiel / framework de compétences aux ressources, activités et sections des différents cours relevant du framework considéré, ce sont les seuls savoirs et savoir-faire (d'un cours) que COMPER permet de relier via leur url aux ressources d'un LMS comme Moodle. L'approche Moodle de la notion de compétence correspond donc à une vision transversale de l'APC alors que celle de COMPER apparaît plus fine mais aussi plus disciplinaire et surtout plus focalisée sur un cours particulier.
- Le tutoriel présentant l'articulation des différents outils COMPER suggère l'utilisation d'un outil Web (*Platform2Ref*) pour associer l'url d'une ressource pédagogique existante aux savoirs et savoirs-faire créés dans le référentiel. En pratique, il nous est apparu très difficile de séparer le volet « Edition d'un référentiel de compétences » du volet « construction / édition des ressources et activités pédagogiques via un LMS ». Si certains enseignants / ingénieurs pédagogiques peuvent mener une approche inductive consistant à d'abord créer le référentiel de compétences avant de produire un jeu de ressources pédagogiques en cohérence avec de référentiel, si d'autres préfèrent mener le processus inverse qui part d'une édition/production de ressources et d'activités, il est clair qu'il serait pertinent de pouvoir passer de l'un à l'autre et qu'idéalement ceci devrait pouvoir être mené via les fonctionnalités du LMS exploité par les enseignants.

Partant de ces constats, nous proposons dans la section suivante des éléments d'un modèle conceptuel pour pouvoir tirer parti dans Moodle des éléments de modélisation pertinents apportés par le métamodèle COMPER.

2.2 Eléments du modèle conceptuel proposé

La Figure 1 illustre une instanciation partielle du modèle que nous proposons pour mettre des enseignants ou un système informatique en situation de recommander aux étudiants des extraits ciblés de ressources pédagogiques lorsque des lacunes sur certains concepts ou skills sont identifiées. D'un point de vue modélisation informatique, il s'agit d'une instanciation du modèle UML qui sera présenté en figure 2.

L'instance en figure 1 montre :

- Deux extraits d'une même ressource Moodle : il s'agit d'une *page* au sens de Moodle sur laquelle des annotations ont été posées (une en bleu et l'autre en vert)
- Des concepts et des skills en bas de figure 2,

- Des liens entre ces concepts / skills et les extraits ciblés le page Moodle au travers d'objets de recommandation (O1 et O2).

Modélisation des concepts et des skills. Une skill est définie comme un savoir-faire très ciblé, pouvant être décrit par l'application d'un verbe d'action (Verb, *e.g.*, « déclarer », « typer », « commenter ») à un concept (Concept, *e.g.*, « variable »). Par exemple : « Déclarer une variable », « Typer une variable » ou « Nommer une variable » (Skills S1 à S3 dans Fig. 1).

Les relations entre skills permettent de structurer les dépendances pédagogiques qui les relient. Par exemple, une skill telle que « Déclarer une variable » peut être composée de skills plus élémentaires comme « Nommer une variable » et « Typer une variable ». Ces relations incluent des notions de :

- Prérequis qui définissent les skills nécessaires avant d'en acquérir une autre.
- Composition où une skill complexe est composée de plusieurs skills élémentaires (Skill S3 est composée de Skill S1 et S2 dans Fig. 1).
  Complexification, représentant une skill de niveau supérieur découlant d'une skill de niveau inférieur.

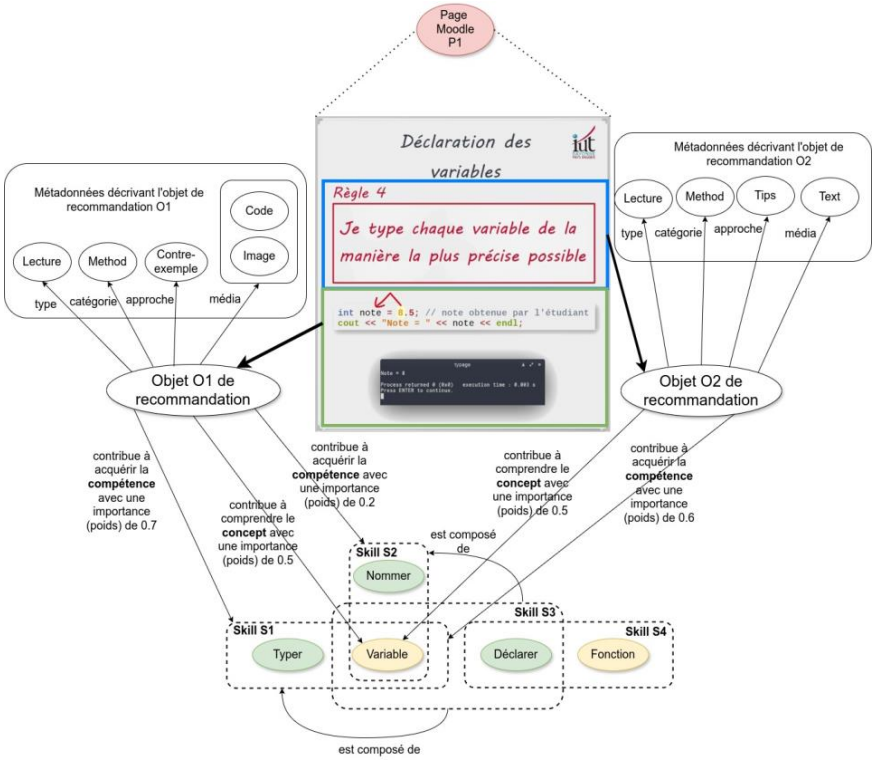


Fig. 1. Exemple d'une ressource pédagogique détaillée à un niveau micro.

Les objets de recommandation et leur structuration. Les objets de recommandation sont des extraits ciblés de ressources Moodle, identifiés pour faciliter la compréhension d'un concept et/ ou l'acquisition d'une skill. Ils sont caractérisés par des métadonnées de nature pédagogique (Fig. 1) :

- Type (Recommendation object type). Modalité d'apprentissage associée à l'objet, par exemple : tutoriel interactif (Tutorial), contenu théorique (Lecture) ou évaluation (Assessment).
- Catégorie (Recommendation object category). Nature de l'apprentissage facilité par l'objet, par exemple : Compréhension d'un concept (*e.g.*, « qu'est-ce qu'une variable ? »), Acquisition d'une méthode (*e.g.*, « comment déclarer une variable ? »).
- Approche (Approach). Nature du contenu présenté, par exemple : Exemple (illustration pratique), Définition (contenu théorique structuré), Solution (résolution d'un problème).
- Média (Media). Support utilisé pour présenter le contenu, par exemple : Texte, Image, Vidéo, Code source, ou une combinaison de ces éléments.

Relations entre concepts, skills et objets de recommandation. Chaque objet de recommandation est associé à un ou plusieurs concepts / skills pondéré par un poids (weight), qui indique la pertinence et l'intérêt relatif de l'objet pour comprendre / développer un concept / une skill donnée. Par exemple, un extrait de code montrant comment déclarer une variable est fortement pertinent pour la skill « Déclarer une variable » mais peut aussi avoir une utilité (moindre) pour des skills connexes comme « Typer une variable » ou « Nommer une variable ». Le modèle permet également de relier les concepts via des relations (Concept relation), offrant une structuration supplémentaire pour modéliser des notions liées (par exemple le concept de « Variable » est une notion de départ pour illustrer ensuite des concepts liés comme les indices de boucles, les compteurs, les accumulateurs, les indicateurs, etc.). Les verbes d'action (Verb) jouent un rôle central dans la définition des skills, en formant une ontologie et un vocabulaire métier spécifique au domaine. Cette structuration permet par exemple d'identifier toutes les skills à maîtriser autour du concept de « Variable » : « nommer une variable », « typer une variable », « commenter une variable », « appliquer une convention de nommage à une variable », « déclarer une variable », « initialiser une variable », etc.

Les objets de recommandation sont ensuite alignés avec ces skills pour servir de support à des activités pédagogiques ciblées et proposer des aides adaptées à des besoins très ciblés d'étudiants en situation d'apprentissage (voir flèches descendantes de O1 et O2 vers les Skills et concepts dans Fig. 1).

2.3 Principes d'exploitation du modèle conceptuel proposé

La modélisation présentée (voir Fig.1) définit une ontologie du domaine métier en associant des concepts et des verbes pour structurer les skills travaillées ou à acquérir. Une skill est ainsi composée d'un verbe (action à réaliser) appliqué à un concept (objet

de l'action). Cette approche offre une structuration du domaine métier qui permet d'explorer les skills selon plusieurs entrées ou angles d'analyse.

- *Une exploration selon l'entrée « Concept »* pour identifier l'ensemble des skills en lien avec un concept. Par exemple, si l'on considère le concept « Variable » dans le domaine de la programmation, les skills qu'un étudiant devra maîtriser sont les suivantes : Typer une variable Skill S1 en bas de la Figure 1), nommer une variable (S2), déclarer une variable (S3), initialiser une variable, etc.
- *Une exploration selon l'entrée « Verbe »* pour explorer les différents concepts sur lesquels ce verbe s'applique. Par exemple, pour le verbe « Déclarer », il est possible de déduire ce qui peut être déclaré dans le cadre de la programmation :
  - Déclarer une variable : introduire une variable dans un programme (Skill S3), Déclarer une fonction : introduire une fonction avec ses paramètres et sa valeur de retour (Skill S4).

  Ainsi, un étudiant qui apprend à « déclarer » pourrait identifier qu'il doit non seulement savoir déclarer des variables, mais aussi des constantes, des fonctions, élargissant ainsi sa maîtrise du verbe « déclarer » dans différents contextes.

D'un point de vue pédagogique, cette modélisation apporte une vision flexible et multi-entrées de l'apprentissage des concepts et des skills, permettant d'aborder les savoir-faire selon des angles pédagogiques différents :

- Angle « Concept » : Quelles sont les actions à maîtriser pour un concept donné ?
- Angle « Verbe » : Sur quels concepts s'applique une action donnée ?

Cette approche permet aussi d'explorer le domaine métier avec une granularité fine, facilitant ainsi :

- L'identification des skills spécifiques à travailler pour comprendre et maîtriser un concept,
- La structuration des séances d'apprentissage autour des relations entre concepts et actions,
- Une meilleure personnalisation de l'apprentissage, en permettant aux étudiants d'accéder aux skills en fonction de leurs besoins spécifiques.

3 Discussion sur les usages et sur la mise en œuvre

Il nous semble qu'enseignants et étudiants peuvent se rejoindre sur l'importance des notions de skill et de concept pour mieux tirer parti de l'approche par compétences dans des pratiques diversifiées et à forte valeur ajoutée :

- Dès lors que des skills ont été attachées à un cours, aux ressources et aux activités Moodle de ce cours, l'apprenant gagne en autonomie dans la manière de naviguer dans ce cours, le graphe de skills devenant une porte d'entrée supplémentaire pour accéder aux contenus Moodle selon une logique qui n'est plus celle d'une organisation en sections et sous-sections mais qui suit une logique de liens sémantiques existant entre skills ou entre des éléments attachés à ces skills.

- Dès lors que l'unité de description d'un cours n'est plus la page dont le titre ou la description expliciterait maladroitement son contenu, mais la skill et tous les concepts qui y sont liés sémantiquement, dès lors que chaque exercice proposé aux étudiants est également annoté sémantiquement selon des skills et autres liens sémantiques caractérisant l'exercice (notions en jeu et skills à mettre en œuvre), tous les contenus d'un cours Moodle deviennent des ressources candidates que l'étudiant peut aller consulter utilement en fonction du contexte de résolution de problème dans lequel il se trouve ; ou encore des ressources qu'un tiers (humain / automatisé) peut suggérer à l'étudiant en fonction d'un diagnostic de ses besoins.

Dans la figure suivante qui décrit notre modèle opérationnel, les concepts issus de COMPER apparaissent en vert, ceux en orange correspondent à des éléments qui sont dans le métamodèle de Moodle, ceux en bleu correspondent aux éléments intégrés dans notre implémentation (sous forme de plugin). Lorsqu'un concept du modèle relève à la fois du métamodèle de COMPER et de celui de Moodle (classes UML en vert et orange), c'est l'interprétation Moodle du concept qui a été choisie car elle est outillée dans Moodle et nous semble mieux cadrer avec une vision transversale de l'APC.

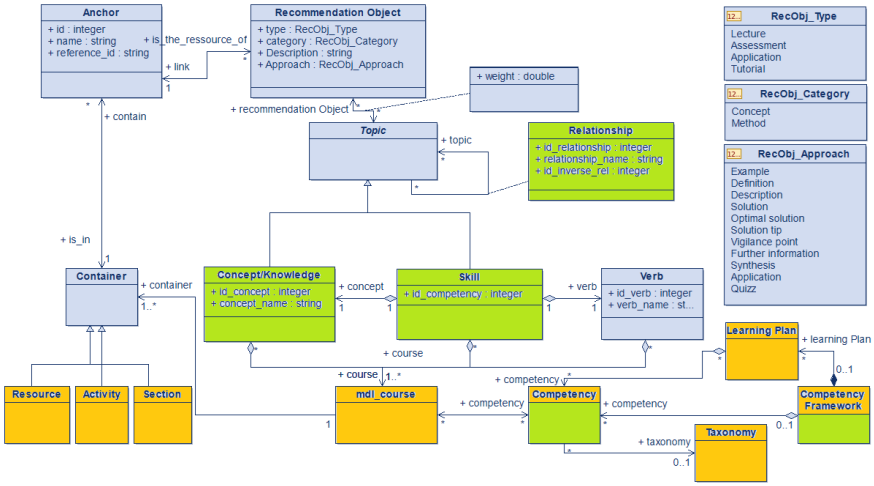


Fig. 2. Modèle proposé (extensions de COMPER et de Moodle).

On remarquera l'apparition des concepts de Container, de Topic et d'Anchor non évoqués jusqu'ici (contrairement au concept d'Objet de recommandation) :

- la classe Container correspond à une entité présente dans le métamodèle de Moodle et implantée sous forme de table dans la Base de données Moodle, quelle qu'en soit la version. Via cette classe, il est possible de créer des ancres sur toute sous-classe d'un conteneur Moodle. Ainsi, toute section Moodle, toute Page Moodle (sous classe de Ressource) devient ancrable, tout forum ou quizz (sous-classes d'Activity) peuvent également faire l'objet d'une pose d'ancre par un enseignant.

- La classe Topic nous permet de définir au niveau de cette super-classe les relations que peuvent entretenir les concepts entre-eux et les skills entre-elles.
- Dans notre système, la classe Anchor permet de garder trace (via notamment un nom choisi par l'enseignant éditeur) d'annotations posés sur un couple (conteneur, Objet de Recommandation). Un même conteneur peut porter différentes ancres ce qui signifie que selon la zone/partie du conteneur ancrée, l'objet de recommandation associé portera des informations spécifiques (cf. skills, concepts et verbes).

Tous les concepts sus-cités ont été implantés dans Moodle comme le montre la vidéo exemple mise en référence [17]. La figure 3 propose un exemple de navigation prenant comme point de départ une page Moodle intitulée *Déclarer ses variables : bonnes pratiques*. Cette page met en avant, via les objets de recommandation qu'elle contient :

- les concepts abordés, tels que variable, constante, et convention de nommage,
- les skills travaillés, comme Déclarer une variable et Nommer une variable (en se limitant à la partie visible du graphe).

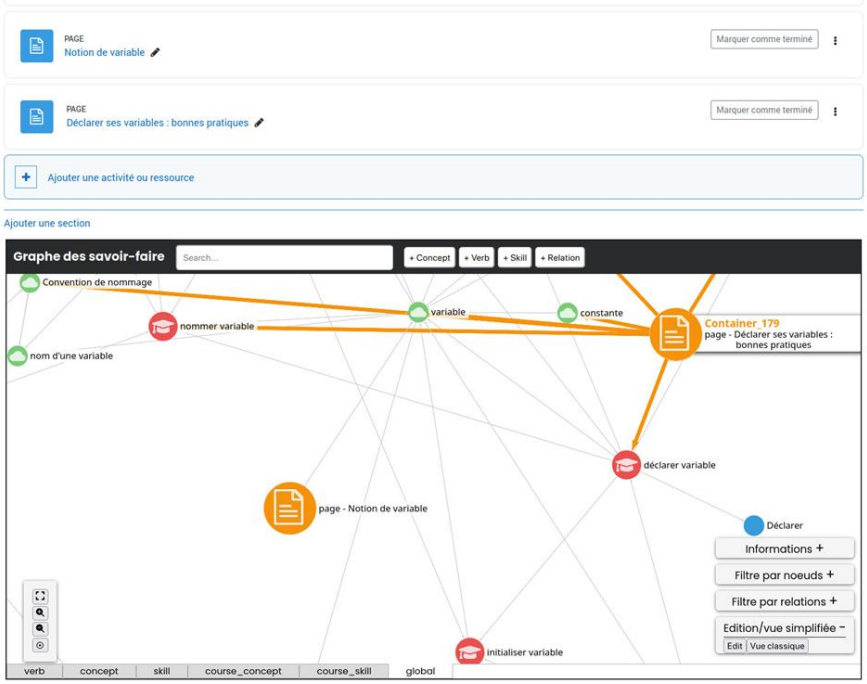


Fig. 3. Copie d'écran d'une partie de graphe de compétences d'un cours de programmation.

Ce graphe met en évidence :

- en orange, les ressources Moodle du cours (deux pages dans cet extrait),
- en vert, les concepts abordés dans le cours,
- en rose, les skills que le cours permet de travailler, sachant qu'une skill est définie comme un verbe d'action (en bleu) appliqué à un concept (en vert).

Lorsqu'un enseignant annote des ressources (par exemple, des pages Moodle), il peut transformer chaque extrait annoté (une définition, un exemple, etc.) en objet de recommandation en reliant l'extrait annoté à des concepts et/ou des skills que l'extrait permet d'aborder. Cette mise en relation peut concerner des concepts et skills déjà existants ou de nouveaux concepts et skills que l'enseignant crée à mesure qu'il annote son cours (cf. boutons d'édition en haut de l'interface).

Ce processus d'annotation permet de construire progressivement une vue détaillée des concepts et des skills couverts dans un cours (ou au sein d'une ressource spécifique du cours). À partir de cette structure, enseignants et étudiants peuvent naviguer dans le cours selon différents points d'entrée.

À ce stade des travaux, nous ne savons pas qualifier la qualité d'un graphe existant sur la base de son organisation interne et nous avons constaté l'intérêt à stabiliser un vocabulaire, notamment lorsqu'un graphe est utilisé dans deux cours Moodle créés par un même enseignant ou par différents enseignants d'un domaine. Pour le domaine particulier de l'apprentissage de la programmation sur lequel nous avons expérimenté l'éditeur de graphes et des contenus Moodle, des ontologies existent [*cf.* http://codeontology.org/] mais restent selon nous peu exploitables. Les liens de synonymie entre skills / compétences sont en l'état un moyen de construire son propre graphe tout en s'inscrivant dans une démarche d'appropriation d'ontologies existantes.

L'implémentation dans le LMS Moodle de fonctionnalités d'édition, de visualisation et d'exploitation de graphes de skills [18] nous a amené à relever plusieurs défis techniques afin de satisfaire aux objectifs suivants :

- En matière d'édition de graphes pour les seuls enseignants éditeurs (au sens de Moodle), offrir des fonctionnalités accessibles à l'échelle du cours Moodle et aussi à l'échelle d'une activité ou d'une ressource particulière : afficher selon le cas tout le graphe du cours ou le seul sous graphe correspondant.
- Offrir un niveau de généricité suffisant pour que les objets de recommandation auxquels sont associés les skills et connaissances puissent pointer une variété de contenus Moodle : le système actuel permet d'opérer sur des contenus d'une session Moodle, d'une page, d'un livre, d'un wiki ou d'un forum Moodle.
- Permettre à plusieurs utilisateurs (éditeurs de cours et étudiants) d'exploiter simultanément les constituants d'un graphe, faciliter notamment l'édition simultanée d'un graphe exploité pour un cours Moodle ou plusieurs cours différents.
- Offrir des mécanismes permettant de passer naturellement de la consultation / édition du graphe à la consultation / édition de la ressource ou activité Moodle sur laquelle s'appliquent les constitutifs du graphe associé à cette ressource / activité via l'objet de recommandation considéré.
- Opérer les extensions envisagées par le seul mécanisme des plugins Moodle.

Le plugin Moodle réalisé [19] comporte deux composantes :

- la première offre une interface de création d'objets de recommandation grâce à un bloc de contenu Moodle qui s'ouvre lorsque l'enseignant est en mode édition de cours. Ce bloc donne à l'enseignant éditeur la possibilité de choisir l'équivalent d'une html parmi celles que contient la ressource ou activité éditée, de la nommer et de lui affecter un objet de recommandation précisant différentes caractéristiques de ce contenu particulier : média, type… l'objet de recommandation est ainsi ancré sur une partie du contenu de l'activité ou de la ressource éditée.
- La seconde offre un cadre qui s'affiche en pied de page et affiche pour l'activité / ressource éditée (respectivement pour le cours édité) un graphe comportant l'ensemble des objets de recommandation définis et pour ceux-ci l'ensemble des compétences et connaissances caractérisant cet objet de recommandation. L'affectation / suppression de compétences ou de connaissances à un objet de recommandation donné se fait en double cliquant sur cet objet de recommandation. La constitution ou la modification du graphe pour manipuler les compétences et les connaissances (ajout de nœuds, ajout de liens d'une relation typée) se fait également via cette interface graphique.

D'un point de vue architecture logicielle, chaque graphe édité dans Moodle est exploité via un logiciel client qui synchronise ce graphe avec le serveur du graphe d'un cours Moodle donné, ce qui permet sans rechargement de page de faire apparaître dans le graphe client tout nœud qu'un autre enseignant éditeur viendrait d'avoir ajouté/modifié via une autre connexion à Moodle. Réciproquement toute modification de graphe faite au niveau du logiciel client est synchronisée automatiquement vers le serveur de graphes.

L'interface permettant d'éditer et d'interagir avec un graphe est réalisée à partir d'une solution `Node.Js` + `Vue.js` + `Sigma.js`. Cote serveur, l'ensemble des graphes est géré/gardé en mémoire via une autre bibliothèque javascript nommée `Graphology`, elle permet de gérer de très gros graphes sans aucune latence pour les utilisateurs.

4 Conclusions et perspectives

Que ce soit en présentiel, en modalité hybride ou en formation à distance, les LMS comme Moodle ont été / se sont imposés comme principaux vecteurs des dispositifs de formation universitaire ; avec leurs qualités indiscutables pour le quotidien des formateurs et des étudiants mais non sans lacunes. Nous avons proposé dans cet article des extensions pour l'APC issues du métamodèle de COMPER qui ne remettent pas en cause les fonctionnalités de gestion des compétences déjà intégrées dans les différents briques de Moodle.

Une version étendue de Moodle intégrant notre plugin a été déployée, elle permet à des enseignants de mener à bien toute l'édition des ressources / activités d'un cours et toute l'édition des compétences / savoirs / savoir-faire dans un seul et même

environnement numérique, ce qui donne une grande liberté méthodologique aux enseignants (démarche inductive commençant par l'expression des compétences, savoirs et savoir-faire ou démarche privilégiant plutôt l'ajout d'un modèle de compétences, de savoirs et de savoir-faire à des cours existants). L'édition sous forme de graphes est disponible à l'échelle d'une activité comme à l'échelle d'un cours, ce qui facilite l'analyse des ressources (taux de couverture des compétences ciblées, diversité des activités proposées) et donc l'édition d'un cours sous l'angle de l'APC.

Les deux ingénieures pédagogiques de l'IUT qui ont contribué aux travaux que présentés exploitent actuellement la version de Moodle ainsi déployée. La mission confiée consiste à développer un cadre méthodologique permettant d'amener progressivement les équipes pédagogiques de deux départements d'IUT à enrichir voire reconstruire leurs supports de cours Moodle pour répondre à l'APC. Il s'agit tout à la fois de mener avec les enseignants les plus mobilisés :

- Un travail d'extraction de leurs connaissances expertes afin qu'ils explicitent les skills / concepts laissés souvent dans l'ombre si l'on s'en tient à la seule lecture de leurs supports Moodle.
- Un travail de confrontation de ces connaissances expertes avec celles des collègues ou celles décrites dans des ontologies tierces.
- Un travail de scénarisation visant à mettre les enseignants en situation de tirer parti des graphes établis, en les proposant aux étudiants comme clés d'appropriation ou d'approfondissement lorsque ces derniers sont en phase de résolution de problème.

Pour ce qui est des perspectives, nous savons que le coût d'étiquetage des ressources et activités selon notre approche compétences est significatif pour l'enseignant. L'assistance à cet étiquetage via des outils à base d'ontologies collaboratives [20] constitue une voie à explorer pour soutenir les enseignants d'un collectif (département d'IUT, Commissions Pédagogiques Nationales par groupe de spécialité de BUT).

Enfin, puisque les travaux relatés dans cet article sont un des volets d'un programme de recherche plus vaste visant à opérer l'APC en situation de résolution de problèmes, nous soulignons l'importance à repositionner les LMS comme Moodle de sorte à ce qu'ils viennent dorénavant en appui de l'apprenant lorsque celui-ci est impliqué dans des tâches authentiques (dans notre cas de programmation) : tout d'abord à l'initiative de l'étudiant qui face à une difficulté pourrait accéder au graphe de skills / compétences de l'exercice qui lui est proposé, naviguer via les différents liens vers des ressources / concepts utiles à la résolution de problème en cours. Mais aussi à l'initiative d'un système informatique (capable d'exploiter les actions de l'étudiant mais aussi les différentes versions du code des programmes qu'il écrit) pour recommander [21], via une boite de message, des ressources Moodle identifiées comme pertinentes que l'étudiant pourra choisir ou pas de consulter. Il nous faudra encore du temps pour que nos prototypes existants de ce système [22] [23] tirent parti des informations stockées dans Moodle et puissent être ensuite validés à l'échelle d'un département d'IUT.

Références

1. Vergnaud, G : La théorie des champs conceptuels, Recherche en didactique des mathématiques, 10, 2/3, 133-170 (1990)
2. LeBoterf G. : L'ingénierie des compétences (2ème édition). Éditions d'organisation, Paris, 445 pages (1999)
3. Gilbert, P. : La compétence : concept nomade, significations fixes. Psychologie du Travail et des Organisations, vol. 12, Issue 2, pp. 67-77. https://www.sciencedirect.com/science/article/abs/pii/S1420253006000227 (2006)
4. Samurçay, R. et Rabardel, P. : Modèles pour l'analyse de l'activité et des compétences, propositions, *in* R. Samurçay et P. Pastré (Éds.), *Recherches en didactique professionnelle,* pp. 163-180, Toulouse : Octarès (2004)
5. Coulet, J.C. : La notion de compétence : un modèle pour décrire, évaluer et développer les compétences », Le travail Humain, Vol. 74, n°1, 1-30 (2011)
6. Messaoui, A. et Pélissier, C. Vers l'approche par compétences : théories et pratiques pour l'enseignement supérieur. Presses des Mines, pp.406, Design Numérique, Pierre-Michel Riccio, 978-2385423353 (2024)
7. LabSET : Laboratoire de soutien aux Synergies Éducation Technologies, Université de Liège. https://www.labset.uliege.be
8. Poumay, M., Tardif, J., Georges, F. : Organiser la formation à partir des compétences. 1ère édition De Boeck Supérieur, Louvain-La-Neuve (2017)
9. Créer un modèle de Plan de formation avec Moodle, https://docs.moodle.org/4x/fr/Plans_de_formation, last accessed 2024/10/18
10. Roussel, C. : Autoévaluation et autorégulation : deux habiletés favorables aux apprentissages en profondeur, Vol. 8, Numéro 1, Bibliothèque et Archives nationales du Qué-bec, ISSN 1927-551X, https://docutheque.uquebec.ca/id/eprint/48 (2019)
11. Guin, N. : Un méta-modèle pour le déploiement de l'APC : le projet COMPER. Anita Messaoui, Chrysta Pelissier. Vers l'approche par compétences : théories et pratiques pour l'enseignement supérieur, Presses des Mines, pp.65-80, 978-2385423353(2024)
12. Van Merriënboer, J. J. G. : *Training complex cognitive skills: A four-component instructional design model for technical training*. Educational Technology (1997)
13. Paquette, G. : An Ontology and a Software Framework for Competency Modeling and Management. *Educational Technology & Society*, 1-21. (2007)
14. Méta-modèle de COMPER : https://comper.fr/sites/default/files/2020-12/L1.1 Méta-modèle référentiels_2.pdf , last accessed 2024/01/12
15. Pierrot, L., Michel, C., Broisin, J., Guin, N., Lefevre, M. et Venant, R. : Assessing COMPER Environment to Support Self-Regulation During AutonomousWork: A Usefulness, Usability and Use Case Study, Open and Inclusive Educational Practice in the Digital World, Springer International Publishing, Cognition and Exploratory Learning in the Digital Age book series (CELDA), pp.193-212, (2022)
16. Sablayrolles L., Guin, N., Lefevre,M. et Broisin., J. : Conception et évaluation d'un processus de personnalisation fondé sur des référentiels de compétences. 10e Conférence sur les Environnements Informatiques pour l'Apprentissage Humain, Fribourg / Virtual, Suisse. pp.226-237 (2021)
17. Vidéo de démonstration du plugin Moodle et de l'éditeur de graphes : https://urls.fr/hagsM1
18. Fontaine B. : Étude du potentiel des technologies basées sur NodeJS pour éditer et exploiter des graphes de compétences plus évolués que ceux du LMS Moodle. Rapport de stage de Master 2 en Informatique, Université de Pau et des pays de l'Adour (2024)
19. Code du plugin Moodle permettant d'annoter des ressources et d'éditer les graphes de skills (créé en sept 2024) : https://github.com/Nodenot/SkillsTagger_App/

20. Spoladore D., Pessot, E. et Trombetta. A. : A novel agile ontology engineering methodology for supporting organizations in collaborative ontology development. Comput. Ind. 151 https://doi.org/10.1016/j.compind.2023.103979 (2023)
21. Roux, L., Nodenot, T., Etcheverry, P., Dagorret, P., Marquesuzaà, C. et Lopistéguy, P. : A learner's behavior model for an E-Learning Hybrid recommender system, Cognition and Exploratory Learning in the Digital Age: 65-83, D. Ifenthaler; D.G. Sampson and P. Isaias Editors, ISBN: 978-3-031-18512-0, Springer (2022)
22. Nodenot T., Deguilhem, M., Cornejo J.M., Nguyen, P., Etcheverry, P. : COSYciels : Regards croisés sur un observatoire de progiciels basés sur l'infrastructure COSY, CNRIUT'2025 - Congrès National de la Recherche des IUT, Bayonne (2025)
23. Erritali M., Grota A., Nodenot, T., Etcheverry, P. : Analyse et Modélisation des Comportements des Étudiants dans un Environnement d'Apprentissage : Approches par Clustering et Réseaux de Neurones récurrents, CNRIUT'2025 - Congrès National de la Recherche des IUT, Bayonne (2025)

Analyse semi-automatique des interactions dans une communauté de pratique : influence du type de rencontre sur la dynamique collective

Sabrin Housni[1], Audrey Kumps[1], Camélia Pissens[2], Karim Boumazguida[1], Gaëtan Temperman[1] et Bruno De Lièvre[1]

[1] Université de Mons, IPN, Mons, Belgique
sabrin.housni@umons.ac.be
audrey.kumps@umons.ac.be
karim.boumazguida@umons.ac.be
gaetan.temperman@umons.ac.be
bruno.delievre@umons.ac.be
[2] Pole Saint-Exupery, Gosselies, Belgique
camelia.pissens@polesaintexupery.be

Résumé. Cette recherche explore l'influence du type de rencontre (partages pédagogiques vs co-construction) sur les interactions au sein d'une communauté de pratique en ligne. Elle s'appuie sur la question suivante : "Le type de rencontre influence-t-il les catégories d'interactions auxquelles sont associés les moments d'intensité de construction des connaissances au sein de la CoP ?". Une méthodologie combinant *ReaderBench* et une analyse de contenu déductive a été mobilisée. Les résultats mettent en évidence des dynamiques distinctes selon le type de séance. Les partages pédagogiques favorisent la transmission d'informations, souvent menée par des experts, tandis que les séances de co-construction suscitent davantage de comportements collaboratifs et socioaffectifs. Cette recherche souligne ainsi la complémentarité des interactions dans les CoP en fonction des objectifs des rencontres : les partages pédagogiques établissent les fondements des connaissances, tandis que les co-constructions permettent leur mise en application.

Mots-clés : Communauté de pratique ; Interactions ; Dynamique collaborative ; ReaderBench ; Analyse de contenu

Abstract. This research explores the influence of the type of meeting (pedagogical sharing vs co-construction) on interactions within an online community of practice. It is based on the following research question: "Does the type of meeting influence the categories of interactions associated with moments of intense knowledge construction within the community of practice?" A methodology combining ReaderBench and a deductive content analysis was employed. The results highlight distinct dynamics depending on the session type. Pedagogical sharing sessions promote the transmission of information, often led by experts,

while co-construction sessions foster more collaborative and socio-affective behaviors. This research thus underscores the complementarity of interactions in communities of practice depending on session goals: pedagogical sharing lays the foundation for knowledge, while co-construction supports its application.

Keywords: Community of practice; Interactions; Collaborative dynamics; ReaderBench; Content analysis

1 Introduction

Les Communautés de Pratique (CoP) constituent un levier pour le développement des compétences professionnelles [1]. En réunissant des participants aux profils divers, elles facilitent l'échange d'expérience et la transformation des pratiques. Au sein des CoP, la qualité des interactions entre les membres revêt une importance particulière, car elle conditionne la richesse des apprentissages [2]. Cependant, ces interactions ne se produisent pas de manière uniforme, et certains moments de la discussion se distinguent par une intensité des échanges. Ces "pics" d'intensité représentent des moments importants de la discussion où la dynamique collective devient fructueuse dans le processus de co-construction des savoirs [3].

Cette dynamique collective repose sur le processus de participation/réification présentée par Wenger [1] comme un des fondements pour comprendre la dynamique d'une CoP. Construire des connaissances ne se résume pas à les acquérir, mais suppose de favoriser leur mobilisation à travers des interactions sociales [2]. Ces échanges d'idées renforcent la capacité des participants à coconstruire de nouveaux savoirs. Par ailleurs, les CoP constituent des environnements propices à l'apprentissage à travers les discussions, les débats ou encore la résolution collaborative de problèmes [4].

Toutefois, le fonctionnement des CoP n'est pas automatique et repose sur des conditions spécifiques [5]. L'intégration ponctuelle de personnes externes à la communauté représente également un enjeu important. Ces contributions externes permettent d'apporter des idées nouvelles et d'éviter que la communauté ne se limite à ressasser des pratiques déjà établies [6].

L'analyse des interactions au sein d'une CoP constitue ainsi une voie pour comprendre son fonctionnement et identifier des traces d'échanges caractéristiques de processus de participation/réification, de co-construction des savoirs, ou d'autres formes d'apprentissage collectif. Ces traces permettent d'évaluer dans quelle mesure la communauté atteint ses objectifs de partage des pratiques, tout en contribuant à la pérennisation de ses dynamiques.

Dans ce contexte, l'objectif de notre recherche consiste à analyser l'influence du type de rencontre au sein d'une CoP en ligne sur les interactions qui s'y déroulent. Cet article se structure en plusieurs parties. D'abord, nous présentons les concepts théoriques liés aux dynamiques d'interaction dans les CoP. Ensuite, la méthodologie décrit le contexte, la question de recherche et le traitement des données. Nous présentons ensuite l'analyse des résultats. Enfin, nos résultats sont discutés au regard de notre ancrage théorique. Pour finir, les conclusions, limites et perspectives de la recherche sont présentées.

2 Ancrage théorique

2.1 Communauté de pratique en ligne

Le concept de CoP, introduit par Wenger [1], trouve son origine dans l'étude des apprentissages sociaux en entreprise [7], avec a pour objectif principal la création et le partage de connaissances [8]. Dans nos travaux antérieurs, nous avons défini les CoP comme un groupe de personnes, aux profils divers, partageant un intérêt commun pour un domaine spécifique. Ces individus se réunissent régulièrement pour échanger sur leurs expériences, partager des ressources et explorer de nouvelles perspectives visant à enrichir leurs connaissances et à améliorer leurs pratiques [9]. Parmi ces membres, plusieurs niveaux de participation coexistent [10-8]. Certains, en tant qu'animateurs, s'investissent activement dans les échanges et la gestion des projets communautaires, tandis que d'autres participent régulièrement aux activités, jouant un rôle moins central.

Charlier et Daele [11] identifient trois variables clés des communautés en ligne: l'entrée, la participation et l'apprentissage. Parmi elles, la participation influence la contribution des membres et la co-construction des savoirs [11]. Elle permet donc une meilleure compréhension des dynamiques de co-construction dans ces environnements.

Par ailleurs, Wenger [1] définit la participation comme le fait de « faire partie de quelque chose avec d'autres » (p. 61), englobant des notions d'appartenance, d'engagement collectif et de reconnaissance mutuelle. La participation dépasse donc la simple présence et constitue un levier essentiel de l'apprentissage collectif au sein des CoP.

Enfin, les conditions de participation renvoient à l'implication effective des membres dans la communauté [11]. Notre recherche s'intéresse à la manière dont les membres y participent à partir de l'analyse des interactions qui s'y déroulent.

2.2 Analyses des interactions dans les communautés

Notre revue de la littérature a mis en évidence de nombreuses typologies d'interactions propres aux environnements en ligne. Étant donné le manque de recherches spécifiques sur les interactions au sein des CoP et la proximité conceptuelle entre les rôles de tuteur et d'animateur [4], les interventions tutorales ont également été prises en compte. Cette démarche a conduit à un regroupement sémantique des différentes catégories d'interactions.

Le point de départ de notre typologie s'appuie sur les travaux de Burnett et Buerkle [12], qui proposent une classification en trois grandes catégories : les comportements non interactifs, les comportements interactifs hostiles et les comportements interactifs collaboratifs. Par souci de concision et parce qu'elles sont les seules interactions identifiées dans notre analyse des résultats, nous avons retenu uniquement celles relevant des comportements interactifs collaboratifs (Tableau 1).

Tableau 1. Typologie des interactions dans les CoP

			Sources
Comportements non spécifiquement orientés vers l'information	Socio-affectif	Comportements neutres : plaisanteries, confidences, ragots, partage d'informations personnelles en dehors de la thématique abordée	[12-13-14-15]
		Comportements humoristiques	[12-14-16]
		Comportements empathiques : déclaration d'encouragement et de soutien, solidarité	[12-14-16-17]
		Réponse à un comportement hostile	[12]
	Organisationnel	Animation, relance, initiation de discussions	[12-13-14]
		Coordination des discussions et des activités	
	Technique	Formulation d'un appel ou d'une demande d'aide technique	[12-14]
		Résolution de problèmes d'ordre purement technique	[12-14-16]
	Rituels d'ouverture et de clôture, salutations		[12-13-14]
Comportements spécifiquement orientés vers l'information	Messages à dimension sociocognitive	Liens vers des ressources externes	[12-13-15-17]
		Mise à jour et annonces d'expériences personnelles	[12-13-17]
		Requêtes (formulées par d'autres participants de la communauté, prises en dehors de la communauté, présentées à la communauté, adressées à une personne spécifiques de la communauté)	[12-13-18]
		Réponse à un requête	
		Projets de groupe dirigés	[12]
	Messages à dimension métacognitive	Expression d'un regard réflexif sur l'apprentissage ou sur le dispositif d'apprentissage	[12-13]

3 Méthodologie

3.1 Contexte de la recherche

La CoP étudiée regroupe divers acteurs du domaine éducatif et de la formation. Elle a été créée dans le cadre du projet Interreg Teach Transition, visant à valoriser un nouveau parcours universitaire « NUMEFA » en Belgique et en France.

L'objectif de la CoP est de favoriser l'échange et la collaboration entre ces acteurs, qui sont invités à partager leurs connaissances, leurs expériences et co-créer autour de thématiques liées aux pratiques technopédagogiques. Chaque année, trois cycles de CoP sont organisés. Chacun se structure autour d'une thématique définie par l'équipe organisatrice. Chaque cycle comprend trois rencontres : une session de partages pédagogiques, durant laquelle des experts sont invités et deux ateliers de co-construction pendant lesquels les membres collaborent pour co-créer des ressources, des outils ou des connaissances. Ces rencontres sont organisées via *Gather.Town*[1].

Notre étude s'est focalisée sur un cycle de CoP comprenant une séance de partage pédagogique et deux ateliers de co-construction. Par ailleurs, une séance

[1] Gather.Town est une plateforme immersive de collaboration à distance offrant un univers 2D interactif.

supplémentaire de partage pédagogique issue d'un autre cycle a été analysée, afin d'examiner les récurrences et divergences entre ces deux types de rencontres.

Le cycle étudié comprend trois séances de 1h30, espacées de deux semaines, réunissant 49 praticiens. 46 praticiens ont participé à la première séance, 31 à la deuxième, et 25 à la dernière. La quatrième séance, appartenant à un autre cycle, a réuni 49 praticiens, dont 44 étaient présents à la première séance.

3.2 Question de recherche

Notre question de recherche est la suivante : "Le type de rencontre influence-t-il les catégories d'interactions auxquelles sont associés les moments d'intensité de construction des connaissances au sein de la CoP ?". Cette question cherche à explorer comment les séances, qu'elles soient de type partages pédagogiques ou co-construction, influencent la nature des interactions et la dynamique des échanges.

3.3 Traitement des données

Notre travail repose sur l'analyse des échanges réalisée avec *ReaderBench* [19]. Utilisant des méthodes de traitement semi-automatique du langage naturel. *ReaderBench* évalue la cohésion textuelle en analysant les réseaux de cohésion entre participants à travers leurs tours de parole et les composants du texte [20-2].L'un de ses principaux atouts réside dans sa capacité à traiter de grandes quantités de données, en représentant chaque tour de parole comme un nœud et en illustrant les relations de cohésion entre ces nœuds. Cette méthode offre ainsi une vision claire de l'interdépendance des participants et de l'intensité de leurs échanges.

Pour construire le corpus analysé, plusieurs étapes ont été suivies. La première étape s'appuie sur des enregistrements vidéo de la CoP réalisés durant l'année académique 2022-2023. Parmi les neuf rencontres organisées, quatre ont été sélectionnées pour cette étude (soit 44,4 % des rencontres). Deux d'entre elles correspondent à des séances de partages pédagogiques, les deux autres sont des séances de co-construction. Le tableau 2 présente le type de séances, le nombre de participants (présents à la séance), le nombre de contributeurs (prenant la parole), ainsi que le taux de contributeurs de chaque séance, calculé en fonction du rapport entre contributeurs et participants.

Tableau 2. Fréquentation des séances analysées

	Rencontre 1	Rencontre 2	Rencontre 3	Rencontre 4
Type	partage	co-construction	co-construction	partage
Contributeurs				
Membres présents	47	31	26	44
Contributeurs	16	28	18	22
Taux de contributeurs	0,34	0,90	0,69	0,50

La deuxième étape consiste à transcrire les vidéos dans un document *Excel*, comportant une syntaxe préétablie[2], où chaque tour de parole est codé avec des balises spécifiques, indiquant le moment de la discussion et l'identité du locuteur. Parallèlement, un codebook est élaboré pour anonymiser les participants, assurant ainsi le RGPD (Figure 1).

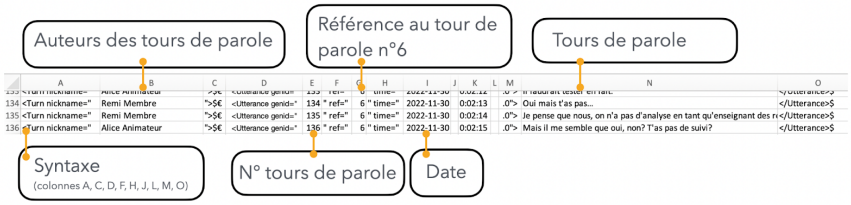


Fig. 1. Transcription des échanges oraux dans *Excel*

Une fois la transcription finalisée dans le fichier *Excel*, une deuxième feuille préétablie permet de générer automatiquement une version du texte à copier dans un éditeur de texte, ici *Visual Studio Code*, afin de produire un document au format XML. Chaque tour de parole est balisé avec des indicateurs spécifiques permettant une lecture structurée des données dans ce logiciel (Figure 2).

[2] Le gabarit est disponible via : https://inspe-sciedu.gricad-pages.univ-grenoble-alpes.fr/rech-educ/tuto-rb-conpa.html

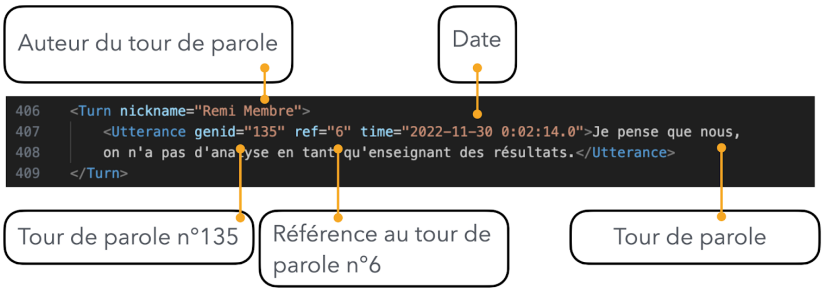


Fig. 2. Transcription des échanges oraux dans *Visual Studio Code*

Le fichier XML est importé dans *ReaderBench* où des scores de contribution par déclaration et contributeur sont disponibles, accompagnés de graphiques illustrant l'intensité des contributions au fil des discussions (Figure 3). Ces visualisations permettent de suivre l'évolution des échanges au sein de la rencontre.

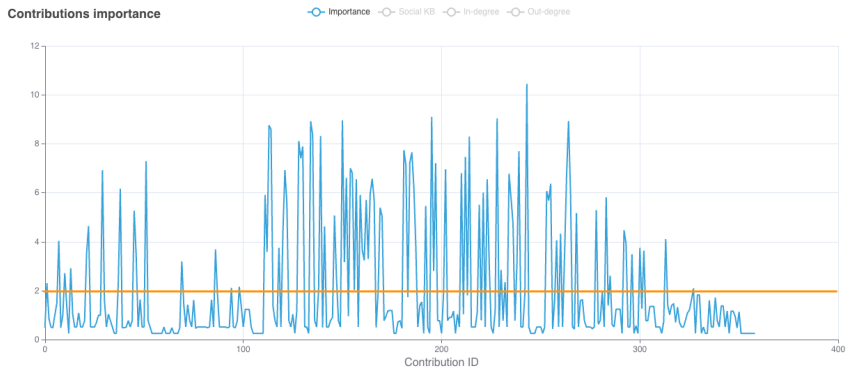


Fig. 3. Graphique de l'importance des contributions au fil des discussions

Finalement, l'analyse des interactions combine les apports de *ReaderBench*, permettant de visualiser les dynamiques conversationnelles à travers des graphiques, et une analyse de contenu déductive selon la méthode de Bardin [21]. Nous structurons cette démarche en trois phases : la préanalyse, l'exploitation du matériel et le traitement des résultats. Les phases de préanalyse et d'exploitation ont conduit à l'élaboration d'une grille de codage, dont les résultats sont présentés dans le cadre théorique, tandis que la méthodologie associée fait l'objet d'un travail distinct, non présenté dans cet article.

La phase de traitement des résultats se concentre sur l'exploration des "pics d'intensité" identifiés dans les graphiques *ReaderBench*. Ces pics, qui correspondent à des moments de discussions particulièrement riches où des échanges fructueux ont lieu [3], sont analysés grâce aux capacités de *ReaderBench* à isoler et zoomer sur des

segments spécifiques. L'outil-est utilisé comme support à l'analyse qualitative, qui reste sous la responsabilité du chercheur [2]. Chaque tour de parole est, dès lors, codé manuellement selon les catégories définies au préalable. Cette méthodologie permet ainsi de relier les graphiques *ReaderBench* à des interactions identifiées dans la littérature.

4 Résultats

4.1 Analyse de l'intensité de la construction des connaissances

La figure 4 montre que les graphiques des "partages pédagogiques" présentent des valeurs maximales plus élevées sur l'axe des abscisses (8 et 12), ce qui rend le graphique plus aéré et facilite l'identification des "pics" d'intensité de construction des connaissances. En revanche, les graphiques des séances de co-construction apparaissent plus denses, avec des valeurs maximales de 3 et 3,5 sur l'axe des abscisses, rendant l'identification des "pics" plus complexe.

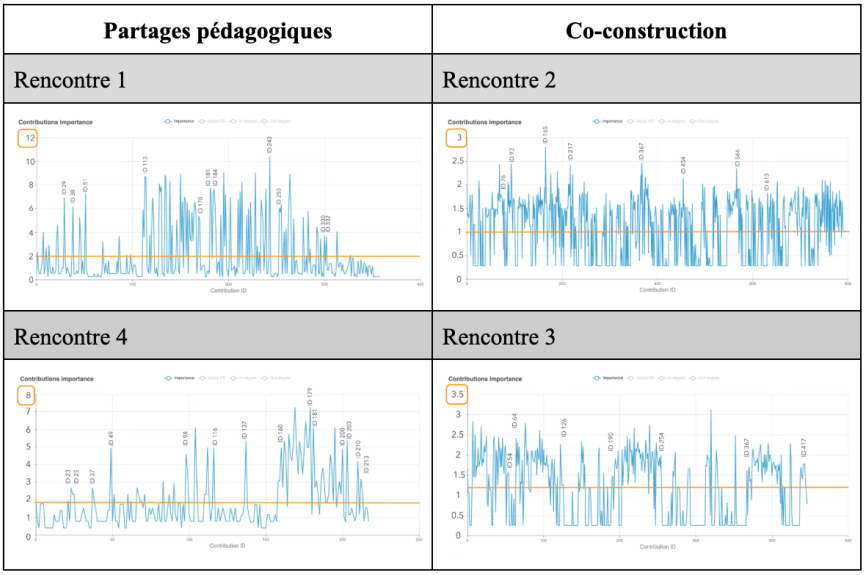


Fig. 4. Comparaison entre les deux types de rencontres

Cette comparaison initiale suggère une première distinction entre les deux types de séances : les "partages pédagogiques" pourraient se caractériser par des prises de parole plus longues, associées à des scores de contribution élevés et facilement détectables sur le graphique des scores de contribution. Cela refléterait une dynamique orientée vers la transmission de savoirs et d'expériences individuelles. À l'inverse, les séances de co-

construction sembleraient présenter des interventions plus courtes, impliquant un plus grand nombre de participants dans des dialogues collectifs.

Dans la section suivante, nous examinons les quatre séances une à une. Nous commençons par les “partages pédagogiques” afin d’explorer les types d’interactions associés aux “pics” d’intensité observés. Cette démarche sera suivie d’une analyse similaire pour les séances de co-construction. Enfin, nous comparerons les deux types de séances.

4.2 Analyse des rencontres de partages pédagogiques

Rencontre 1

La figure 5, illustrant l’intensité des tours de parole au fil de la séance, révèle plusieurs moments clés.

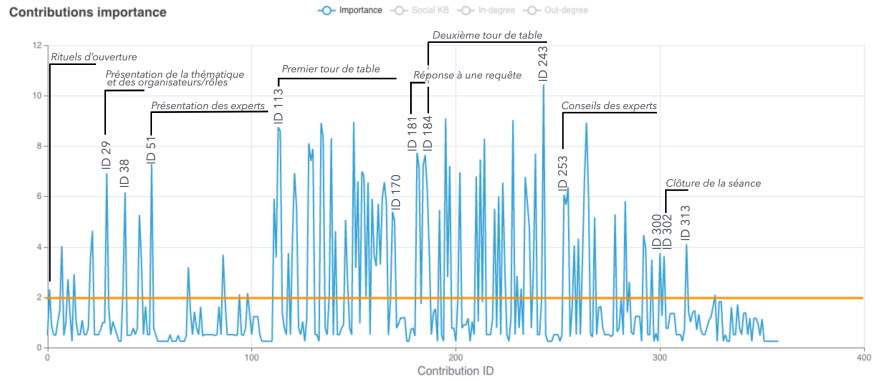


Fig. 5. Graphique de la construction des connaissances de la rencontre 1

Dès le début, avant le tour de parole 29 (ID 29), cinq pics d’intensité associés à des rituels d’ouverture [14] sont observés. Cette phase se distingue par des échanges axés sur l’organisation, la clarification des objectifs et l’accueil des participants.

La première hausse d’intensité entre l’ID 29 et l’ID 38, correspond à la présentation de la thématique abordée en séance et de l’équipe organisatrice. Les comportements sont interactifs collaboratifs, mais non orientés vers l’information [12]. Ce passage témoigne d’une dynamique de coordination, où les animateurs structurent le déroulement des activités [13-14].

Un pic d’intensité, observable à l’ID 51, correspond à la présentation d’un expert. Ce tour de parole est marqué par sur le partage d’informations personnelles non orientées vers l’information [12-13].

La section de l’ID 113 à l’ID 170, associé à un premier tour de table, se caractérise par une alternance de pics et de légères baisses d’intensité, reflétant les tours de parole successifs des membres et des experts. Les experts adoptent des comportements orientés vers l’information, notamment des annonces ou des réponses à des requêtes

[12]. On observe également des explications des experts sur la thématique traitée [17], qui enrichissent la discussion collective. Les animateurs, quant à eux, jouent un rôle d'animation en relançant les débats et en incitant la prise de parole [13]. Dans une moindre mesure, une requête, réalisée par un membre, est adressée à un membre spécifique de la communauté, à savoir un expert [12].

Un pic d'intensité se situe au niveau de l'ID 181 et de l'ID 184. Ces tours de parole mettent en évidence une réponse directe d'un expert à la requête d'un membre [12]. Bien que la demande initiale ne semble pas générer un score de contribution particulièrement "intense", elle suscite néanmoins une réponse de l'expert, ce qui souligne de l'importance des sollicitations individuelles dans la dynamique globale. Ce type de séquence illustre comment des tours de parole apparemment "faibles" peuvent jouer un rôle dans l'enrichissement des échanges.

Un second tour de table commence à l'ID 184, introduit par un témoignage d'un membre de la CoP, associé à une mise à jour personnelle [12]. Ce tour de parole suscite un nouveau pic d'intensité. Les échanges suivent un schéma similaire à celui du premier tour de table : les experts partagent des explications et répondent aux sollicitations des animateurs [12-17] qui interviennent ponctuellement pour animer le débat [13]. De plus, les membres réagissent aux tours de parole des experts en partageant leurs expériences personnelles [12]. Ce moment se distingue par l'apparition de requêtes adressées aux experts par des membres [12].

L'intensité des tours de parole augmente à nouveau entre l'ID 253 et l'ID 300, une phase marquée par les conseils des experts. Ces tours de parole consistent principalement en des apports d'informations et des conseils [12-17]. Nous observons que ces conseils sont majoritairement amenés par les experts, avec quelques contributions ponctuelles des membres.

Enfin, la dernière phase (ID 302 à ID 313), concernant la clôture de la séance, est marquée par une baisse de l'intensité. Les tours de parole sont axés sur la coordination des séances suivantes [13].

Rencontre 4

Le figure 6 illustre l'intensité des contributions de la rencontre 4. En raison d'un problème technique avec l'un des experts, les animateurs ont dû adapter le déroulement de la séance.

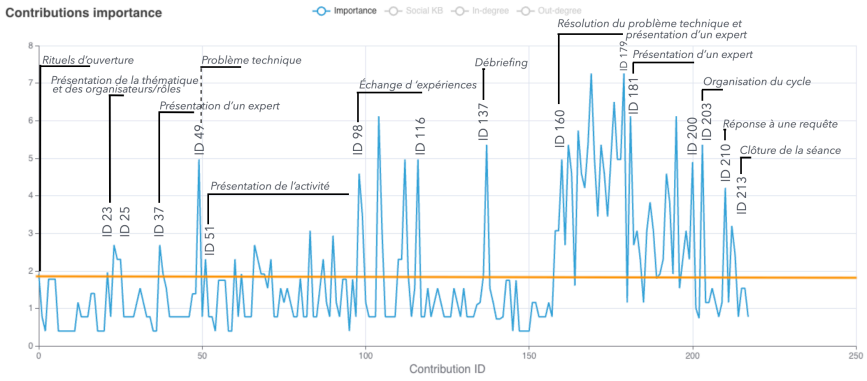


Fig. 6. Graphique de la construction des connaissances de la rencontre 4

Dès le début, avant l'ID 23, on observe des rituels d'ouverture [14]. Entre les ID 23 et 25, les animateurs présentent la CoP et les organisateurs de la séance. Ces contributions s'inscrivent dans la coordination des discussions et des activités [13].

Un pic d'intensité (ID 37) correspond à l'introduction des experts par les animateurs. Ce moment est marqué par des comportements centrés sur le partage d'informations personnelles [12-13].

À l'ID 49, un problème technique est signalé. À ce moment, les animateurs adoptent des comportements centrés sur la gestion du problème technique [14]. La dynamique de la séance est temporairement interrompue.

Pour compenser ce moment d'attente, les animateurs relancent la discussion (ID 51) par l'introduction d'une animation et l'initiation du dialogue [13]. Cette phase se distingue par des comportements organisationnels [12-14]. Ce moment marque le début d'une phase d'échanges plus participatifs.

La période de l'ID 98 à l'ID 116 est marquée par des hausses d'intensité des contributions. Les animateurs ouvrent un moment d'échange d'expériences en lien avec la thématique en incitant les membres à réagir. Les animateurs adoptent des comportements de relances et d'initiations au discours [13]. Les membres, en réponse, partagent des retours d'expériences personnels [12-13].

À l'ID 137, un moment de régulation est observé. L'animateur fait un débriefing sur les contributions précédentes, ce qui se traduit par un comportement d'animation [12-13].

Une hausse de l'intensité survient entre l'ID 160 et l'ID 179, marquée par le retour de l'experte ayant rencontré un problème technique. L'experte est invitée à se présenter et à se positionner par rapport aux retours d'expériences des membres de la CoP. De manière similaire, l'ID 181 à l'ID 200 est marqué par la présentation du second expert. Ces moments se distinguent par des comportements orientés vers l'information, avec des annonces et des explications thématiques [12-17].

À l'ID 203, un pic d'intensité est identifié lorsque l'animateur fournit une explication sur l'organisation du cycle de CoP. Cette intervention introduit un nouvel élément d'information dans la séance. Les comportements adoptés par l'animateur visent la

coordination du cycle, permettant de clarifier des aspects organisationnels pour le bon fonctionnement de la communauté [12-13].

Un nouveau pic d'intensité est observable à l'ID 210, où un expert répond à une requête directe d'un membre. Cette interaction illustre à nouveau comment ces requêtes individuelles peuvent susciter des réponses de la part des experts favorisant des pics d'intensité dans la dynamique globale. Ce type d'échange renforce l'importance des requêtes adressées à une personne spécifique de la communauté dans la construction des connaissances collectives [12].

Enfin, la séance se clôture avec l'ID 213, où l'animateur conclut la séance. Ce moment est marqué par des comportements centrés sur la coordination des activités [13-14]. Les animateurs assurent une transition vers la clôture, marquant la fin des échanges.

4.3 Analyse des rencontres de co-construction

Rencontre 2

La figure 7 illustre l'intensité des contributions au fil de la rencontre 2.

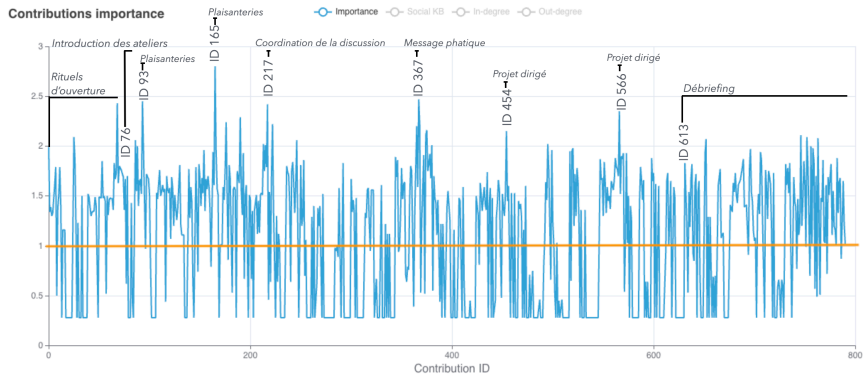


Fig. 7. Graphique de la construction des connaissances de la rencontre 2

Comme pour les autres séances, cette rencontre commence par les rituels d'ouverture réalisés par les animateurs [14], avant l'ID 76. Cette phase est dédiée à la présentation des objectifs et à la coordination de la rencontre [13].

À l'ID 76, la dynamique de la rencontre change avec l'introduction des cinq ateliers. Chaque atelier, composé de 4 à 7 contributeurs, débute par la même phase. Un animateur introduit un projet de groupe dirigé [12], suivi par un travail collaboratif guidé par l'ensemble des membres. Les contributions sont orientées vers la collaboration. Des comportements orientés vers l'information dominent ces ateliers [12].

Un aspect intéressant de cette phase est l'apparition de nouvelles formes d'interactions. Parmi celles-ci, des comportements socioaffectifs sont observés, tels que

des plaisanteries et des échanges informels, en lien avec la dimension socioaffective des échanges [12-14]. Une autre nouvelle forme d'interaction se manifeste par des messages à dimension métacognitive [13]. Les membres expriment un regard réflexif sur le dispositif d'apprentissage, projetant leurs propres pratiques et expériences personnelles dans le contexte des ateliers.

Tout au long des ateliers, des pics isolés d'intensité sont identifiés, notamment à l'ID 93 et à l'ID 165, où des plaisanteries émergent entre les participants [12]. À l'ID 217, un pic d'intensité est associé à une intervention de coordination de la discussion, tandis qu'à l'ID 367, un message phatique de remerciement est relevé [13]. Par ailleurs, des contributions concernant des projets dirigés génèrent également des pics d'intensité (ID 454, ID 566).

Enfin, l'allure du graphique à partir de l'ID 613 montre que la phase de débriefing est fructueuse en termes de construction de connaissances, certainement parce qu'elle a consisté en un bilan général de la séance.

Rencontre 3

La figure 8 illustre l'intensité des contributions au fil de la rencontre 3.

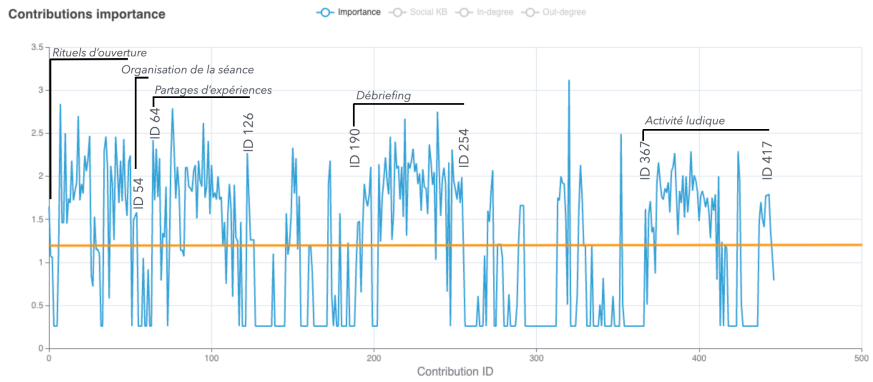


Fig. 8. Graphique de la construction des connaissances de la rencontre 3

La séance commence par un rituel d'ouverture, observé avant l'ID 54, où les animateurs présentent les objectifs de la rencontre. Ce moment est centré sur la coordination des discussions [12-13-14].

Entre l'ID 64 et l'ID 126, les animateurs introduisent une activité d'animation pour répartir les membres dans différents ateliers. Les contributions sont majoritairement centrées sur des mises à jour personnelles et des partages d'expériences [12-13-17].

Entre l'ID 191 et l'ID 254, un débriefing avec l'ensemble des membres est proposé. Ce temps est marqué par des échanges entre les animateurs et les membres. Les animateurs partagent des récapitulatifs des projets dirigés réalisés en amont de cette séance et de la précédente (Rencontre 2). Les membres apportent leurs retours et réflexions, générant ainsi une dynamique de réflexion collective [13].

À partir de l'ID 367, un jeu de fin de cycle est introduit. Cette activité est caractérisée par des comportements socioaffectifs, où l'humour et les échanges informels dominent [12-14]. Les membres interagissent de manière informelle, marquant une transition vers la clôture de la séance et du cycle.

5 Discussion

5.1 À propos des rencontres de "partages pédagogiques"

Les partages pédagogiques se distinguent par une structure d'interactions relativement stable entre les deux rencontres. Les rituels d'ouverture [14] marquent le début de la séance, favorisant la mise en place de l'organisation de la rencontre. Cette phase se caractérise par des comportements interactifs collaboratifs non orientés vers l'information [12], centrés sur la coordination de l'activité [13].

Au cœur des séances, l'introduction d'experts est centrale, comme en témoigne l'augmentation de l'intensité des échanges lors de leur présentation (ID 51, rencontre 1 ; ID 38, ID 181, rencontre 4). Ces moments sont caractérisés par des comportements collaboratifs orientés vers l'information, notamment à travers des annonces et des explications thématiques [12-17]. Les interventions des experts fournissent des informations pratiques et des conseils immédiatement exploitables par les membres de la CoP. Ces derniers participent activement à cette dynamique en sollicitant directement les experts par des requêtes ciblées [12]. Cette diversité de perspectives contribue à enrichir les discussions collectives et à élargir les points de vue sur les sujets abordés. Selon Daele [6], l'apport temporaire d'experts externes constitue un vecteur essentiel d'apprentissage et de pérennisation pour la communauté. En introduisant des perspectives nouvelles, ces experts permettent à la CoP d'éviter de ressasser constamment les mêmes idées, favorisant ainsi son ouverture et son renouvellement.

La phase de clôture marque la fin progressive de la séance. Elle est caractérisée par une diminution de l'intensité des contributions (ID 302 à ID 313, rencontre 1 ; ID 213, rencontre 4). Ce moment est orchestré par les animateurs, qui assurent la coordination des activités futures et la structuration des étapes à venir [13]. Les comportements interactifs sont principalement non orientés vers l'information [12] : remerciements, bilan de la séance et précisions sur les activités à entreprendre par la suite.

Finalement, les interactions des séances de partages pédagogiques suivent une organisation cyclique : ouverture, intervention des experts, séances de questions-réponses et clôture. Cette structure se retrouve dans les deux séances analysées (rencontres 1 et 4). Malgré des imprévus, comme un problème technique au début de la rencontre 4, la continuité des dynamiques observées entre les deux séances met en évidence une certaine récurrence dans les interactions.

5.2 À propos des rencontres de "co-construction"

Les séances de co-construction se distinguent des séances de partages pédagogiques par des dynamiques plus collaboratives. Si les ouvertures suivent un schéma similaire, les

ateliers collaboratifs, les échanges socioaffectifs et les débriefings collectifs marquent les principales différences.

Le cœur des séances de co-construction se situe dans les ateliers collaboratifs (ID 76, rencontre 2 ; ID 64, rencontre 3), centrés sur des projets de groupes dirigés [12], qui favorisent une participation équilibrée des membres. Ces moments se caractérisent par des échanges collectifs où chaque participant apporte une contribution à la co-construction des connaissances. Ces interactions permettent une confrontation d'idées et de pratiques variées entre les membres de la communauté, donnant lieu à des débats enrichis par la diversité des approches et des argumentations. Comme l'indique Daele [6], l'argumentation occupe une place importante dans les processus d'apprentissage au sein des communautés, agissant comme un levier pour le développement des savoirs collectifs.

Les interactions socioaffectives (ID 93, ID 165, rencontre 2 ; ID 367, rencontre 3) sont plus marquées dans ces séances. Les moments de plaisanterie, d'encouragement et d'échanges informels [12-13] participent au climat socioaffectif au sein des groupes. Ces comportements, absents ou moins fréquents dans les séances de partages pédagogiques, pourraient renforcer l'engagement des membres et créer une dynamique plus participative, propice à l'apprentissage [6].

Enfin, le débriefing (ID 613, rencontre 2 ; ID 191, rencontre 3) clôture les séances, permettant une réflexion collective sur les apprentissages réalisés et les actions à venir. Contrairement à la clôture des séances de partages pédagogiques, qui se caractérise par une baisse de l'intensité des échanges, le débriefing des séances de co-construction génère une dynamique plus soutenue.

En résumé, les séances de co-construction se caractérisent par des ateliers participatifs et un débriefing fructueux, contrastant avec les échanges davantage centrés sur les interventions des experts observés dans les séances de partages pédagogiques. Ces interactions participent au processus de participation et de réification, au cours duquel les membres partagent leurs pratiques professionnelles et formalisent les apprentissages issus de leur engagement dans la communauté [1-6].

5.3 Conclusions, limites et perspectives

Cette recherche met en évidence la diversité des interactions dans les CoP selon les objectifs des séances. Les séances de partages pédagogiques privilégient la transmission d'informations, souvent par les experts, tandis que les séances de co-construction encouragent un processus de participation/réification et des comportements socioaffectifs. Ces séances se complètent : les premières établissent les bases des connaissances, et les secondes offrent un espace de mise en application.

L'un des principaux apports de cette étude est la méthodologie d'analyse. L'outil d'analyse semi-automatique des contributions permet d'identifier des tendances et des moments d'intensité dans les échanges. Cette approche, alliant analyse qualitative et semi-automatique, facilite la compréhension des comportements interactifs et offre des indications pratiques pour la gestion des CoP. L'analyse montre que la distinction entre les séances de partage et de co-construction influence la nature des interactions.

Toutefois, la recherche présente certaines limites. Le corpus d'analyse n'est constitué que de quatre séances, dont une provenant d'un cycle distinct, ce qui peut fausser certaines conclusions. En outre, ces quatre séances représentent une fraction des neuf rencontres complètes de la CoP, ce qui peut limiter la compréhension des dynamiques sur le long terme. Enfin, l'utilisation d'outils d'analyse semi-automatique pose également des défis, notamment en raison d'incohérences relevées dans l'attribution des scores pour des tours de parole similaires. Bien que ces outils aient été validés dans des travaux précédents [20], ces limitations invitent à une interprétation prudente des résultats.

Plusieurs perspectives de recherche émergent. l'analyse d'une saison complète de CoP, comprenant 9 séances annuelles, permettrait d'observer des dynamiques plus globales et d'étudier l'évolution des interactions sur une période prolongée. Une analyse longitudinale par membre et par rôle permettrait de comprendre les contributions différenciées et leur influence sur la dynamique des interactions et l'évolution des pratiques. Par ailleurs, une analyse quantitative des interactions permettrait d'affiner les résultats et d'offrir des perspectives d'interprétation plus robustes. Il nous semble intéressant de suivre la participation des membres au fil des séances pour comprendre comment la composition du groupe (nombre de participant, rôles attribués, ancienneté, *etc*.) influencerait les échanges. Enfin, la présente recherche se concentre sur les interactions orales lors des rencontres de la CoP, mais un autre canal de communication, le chat, pourrait également jouer un rôle dans la dynamique collective. L'étude des échanges écrits, en complément des échanges oraux, permettrait d'examiner leur effet sur la dynamique des interactions.

Ces perspectives de recherche ouvrent de nouvelles voies pour approfondir l'analyse des dynamiques au sein des CoP et aider à l'élaboration de dispositifs de formation plus adaptés aux besoins de ses membres.

Références

1. Wenger, E.: *Communities of practice: Learning, meaning, and identity*. Cambridge University Press (1998).
2. Dessus, P., Dascalu, M., Mandran, N., Gutu-Robu G.S., Dormoy-Fournier, C., Ruseti, *S.S. : L'analyse sémantique automatique pour étudier les discussions visant la construction collaborative de connaissances*. In : Albero, B., Thievenaz, J. (eds.) Enquêter dans les métiers de l'humain. *Traité de méthodologie de la recherche en sciences de l'éducation et de la formation,* vol.2, pp.575-589. Raison et Passions (2022).
3. Dessus, P., Simon, J.-P., Dascalu, M., Trausan-Matu, S. : *ReaderBench : Un outil d'aide à l'analyse de discussions philosophiques*. In : Simon, J.-P., Tozzi, M. (eds) Paroles de philosophes en herbe : regards croisés de chercheurs sur une discussion sur la justice en CM2, pp.151-171. UGA Éditions (2017).
4. Nezet, M. : *Les communautés de pratique en entreprise sous l'angle de leur animation : analyse et enjeux*. Ecole Doctorale Sciences de l'Homme et de la Société (2015).
5. Tremblay, D.-G. : *Les communautés de pratique : quels sont les facteurs de succès?*. Revue internationale sur le travail et la société 3(2), 52-79 (2005).

6. Daele, A. : *Les communautés d'apprentissages Conditions d'émergence et impacts pour la formation des enseignants*. Pôlecast (2023).
7. Lièvre, P., Bonnet, E., Laroche, N. : Etienne Wenger. *Communauté de pratique et théorie sociale de l'apprentissage*. In : Burger-Helmechen, T., Hussler, C., Cohendet, P. (dir.). Les grands auteurs en management de l'innovation et de la créativité, pp.427-447. EMS éditions (2016).
8. Bedoya Gonzalez, J. R., Cardona, B., Olrando, M., Montaya, V., León F.: Creación de una comunidad de práctica para la formación de docentes en la integración de las TIC a los procesos de aprendizaje y enseñanza de lenguas extranjeras. *Ikala* 23(1), 121-139 (2018).
9. Housni, S., Kumps, A., Marchal, P., Boumazguida, K., Temperman G., De Lièvre, B. : *Mesure de la présence dans les communautés en ligne. Conception et validation psychométrique d'une échelle de mesure de la présence à distance dans les communautés de pratique en ligne.* Education et formation (sous presse).
10. Wenger, E., McDermott, R., Snyder, W. M.: *Cultivating Communities of Practice : A Guide to Managing Knowledge.* Harvard Business School Press, Boston (2002).
11. Charlier, B., Daele, A.: Chapitre 9 – Pourquoi les communautés d'enseignants aujourd'hui ?. In : Daele , A., Charlier, B. (dir.) *Comprendre les communautés d'enseignants. Pratiques et recherches*. L'Harmattan, Paris (2006).
12. Burnett, G., Buerkle, H.: Information Exchange in Virtual Communities: a Comparative Study. *Journal of Computer-Mediated Communication,* 9(2), (2004).
13. Quentin, I., Condé, J.: Les rôles endossés par les apprenants les plus actifs dans un forum de MOOC : le cas d'ItyPA 2. *Distances et médiations des savoirs*, 15, (2016).
14. Quintin, J.-J. : *Accompagnement tutoral d'une formation collective via Internet - Analyse des effets de cinq modalités d'intervention tutorale sur l'apprentissage en groupes restreints. Université de Mons*, Université Stendhal-Grenoble 3 (2008).
15. Renaud, L., Caron-Bouchard, M., Gaudreault-Perron, J., Gayraud, H. : Communauté de pratique dans le domaine de la promotion de la santé : analyse du sentiment d'appartenance et des pratiques de leadership. *Communiquer*, 19, 29-45 (2017).
16. Prost, M., Cahour, B. et Détienne, F.: Le partage d'émotions et de connaissances sur la pratique : dynamiques des échanges dans les communautés de pratique virtuelles de professionnels. *Le travail humain* ,77(2), 177-202 (2014).
17. Klemm, P., Hurst, M., Dearholt, S. L., & Trone, S. R.: Gender differences on Internet cancer support groups. *Computers in nursing*, 17(2), 65-72 (1999).
18. Savicki, V., Lingenfelter, D., & Kelley, M.: Gender language style and group composition in Internet discussion groups. *Journal of Computer-Mediated Communication*, 2(3), (1996).
19. Dascalu, M.: *L'analyse de la complexité du discours et du texte pour apprendre et collaborer. Université de Grenoble*. Universitatea politehnica (2013).
20. Dascalu, M., Trausan-Matu, S., McNamara, D. S., Dessus, P. : ReaderBench: Automated evaluation of collaboration based on cohesion and dialogism. *International journal of computer-supported collaborative learning*, 10, 395-423 (2015).
21. Bardin, L : *L'analyse de contenu*. Presses Universitaires de France, Paris (1977).

Conception et évaluation : relever les défis de la recherche collaborative menée sur les jeux éducatifs

Mariem Jaouadi[1], Estelle Prior[1], Eric Sanchez[1], Nadine Mandran[2]

[1] Université de Genève, LIP/TECFA, Genève, Suisse
{mariem.jaouadi, estelle.prior, eric.sanchez}@unige.ch
[2] Université de Grenoble Alpes, LIG, Grenoble, France
nadine.mandran@univ-grenoble-alpes.fr

Résumé. Les recherches orientées par la conception (ROC) sur les jeux éducatifs engagent des équipes pluridisciplinaires confrontées à de nombreux défis. Cet article identifie six défis, issus de la littérature et d'un travail empirique mené sur un projet de conception et d'évaluation d'un jeu d'apprentissage. Ces défis ont été discutés lors d'un atelier-jeu réunissant des experts du domaine, qui ont formulé des préconisations accompagnant la conduite de projets en ROC. Si ces préconisations valident la pertinence des défis identifiés et offrent des pistes pour initier des projets, elles nécessitent un approfondissement pour répondre aux besoins des acteurs expérimentés. Afin de relever les défis interdisciplinaires de la conception et de l'évaluation de jeux en ROC, l'article souligne enfin la nécessité de concevoir des cadres méthodologiques pour caractériser trois questions centrales de la ROC : le partage des savoirs, la co-problématisation des questions de recherche et la co-élaboration des indicateurs.

Mots-clés : Recherche orientée par la conception, jeu éducatif, spécification des traces, co-problématisation, partage de savoirs.

Abstract. Design-Based Research (DBR) on learning games involves multidisciplinary teams facing numerous challenges. This article identifies six challenges derived from the literature and an empirical study conducted on a game design project. These challenges were discussed during a game-based workshop with domain experts, who proposed recommendations to support the management of DBR projects. While these recommendations validate the relevance of the identified challenges and provide guidelines for initiating projects, they require further refinement to address the needs of experienced stakeholders. To tackle the interdisciplinary challenges of game design in DBR, the article emphasizes the need to develop methodological frameworks to address three central questions: knowledge sharing, co-problematization of research questions, and co-development of indicators.

Keywords: Design-based research, game-based learning, trace specification, co-problematization, knowledge sharing.

1 Introduction

Les jeux éducatifs représentent des objets complexes. Leur conception est un processus pluridisciplinaire qui implique une méthode favorisant collaboration et partage des cadres théoriques entre les acteurs impliqués.

Dans ce cadre, la recherche orientée par la conception (ROC) [1], constitue une méthode adaptée pour concevoir et évaluer des jeux éducatifs. Cette approche itérative, collaborative, contributive et conduite en contexte écologique, permet de dissoudre le dualisme entre théorie et pratique [2]. Dans cet article, nous défendons une vision où collaborer repose sur une dynamique d'interaction mutuellement enrichissante entre praticiens et chercheurs [3] impliquant négociation, partage et co-construction des pratiques et des savoirs.

La ROC inclut des cycles de conception et d'évaluation menés en parallèle. Toutefois, la gestion des interactions présente une difficulté particulière [4]. Ces enjeux sont partiellement abordés dans la littérature mais leur investigation empirique reste peu documentée. Il est également nécessaire d'explorer des solutions pour surmonter les défis des équipes impliquées dans un projet de conception de jeu éducatif mené en ROC. Ce double besoin alimente notre questionnement : *Quels sont les défis auxquels peuvent être confrontées les équipes pluridisciplinaires lors d'une conception de jeux éducatifs menée en ROC ? Comment ces défis se manifestent-ils concrètement sur le terrain ? Et quelles sont les préconisations des experts pour relever ces défis ?*

Dans cet article, nous explorons les défis issus de la littérature sur la ROC, en nous concentrant sur son application aux jeux éducatifs et les défis identifiés lors de la conception et de l'évaluation d'un jeu d'apprentissage. Ce premier terrain a permis de relier les défis théoriques à des obstacles concrets rencontrés sur le terrain, et de faire émerger de nouveaux défis spécifiques au contexte étudié. Nous avons ensuite confronté ces défis à des experts lors d'un atelier-jeu organisé à la conférence EIAH2023. Cet atelier, original par son format ludique, visait à mettre en situation ces défis pour permettre à ces experts d'explorer des solutions pour les relever et en tirer des préconisations.

La démarche de recherche que nous suivons distingue parfois les phases de conception et d'évaluation. Ce choix est motivé par la nécessité d'alimenter deux travaux de thèses complémentaires menées par les deux premières autrices de cet article et supervisées par les auteurs 3 et 4. Le premier sujet de thèse porte sur le soutien à la collaboration dans la conception de jeux en ROC, en se centrant sur le partage de savoirs entre les acteurs impliqués. Il se focalise sur trois phases clés : démarrage de la conception, opérationnalisation des objectifs pédagogiques et articulation des éléments ludiques et pédagogiques. Le second sujet de thèse explore les processus d'évaluation en ROC, et plus spécifiquement la production des traces numériques d'interaction, définies en collaboration entre praticiens et chercheurs. En effet, la définition de ces traces constitue un point clé de l'évaluation en reliant les questions de recherche aux investigations empiriques.

La littérature propose plusieurs cadres pour la ROC ou la conception de jeux éducatifs, mais peu d'approches les articulent explicitement. Ce travail met en lumière des défis parfois identifiés dans la littérature, mais encore difficiles à lever, peu explorés empiriquement ou insuffisamment outillés. Ainsi, notre contribution s'articule autour

de trois étapes, 1) une revue de littérature sur les défis théoriques liés à la conception et l'évaluation des jeux éducatifs en ROC, 2) une analyse empirique de ces défis dans le cadre d'un terrain d'étude, projet TSADK, et 3) une confrontation de ces défis à un panel d'experts lors d'un atelier-jeu.

Ce dernier, conçu comme une expérience immersive, vise à favoriser le partage des expertises autour des défis identifiés, stimuler la créativité en abordant les défis de façon ludique et encourager la collaboration. Par cette approche, notre contribution enrichit la compréhension des défis de la conception de jeux menées en ROC (volet heuristique), et propose le jeu comme un outil pour faciliter le dialogue et la collaboration entre experts (volet pragmatique).

La figure 1 décrit la manière dont notre recherche s'est articulée.

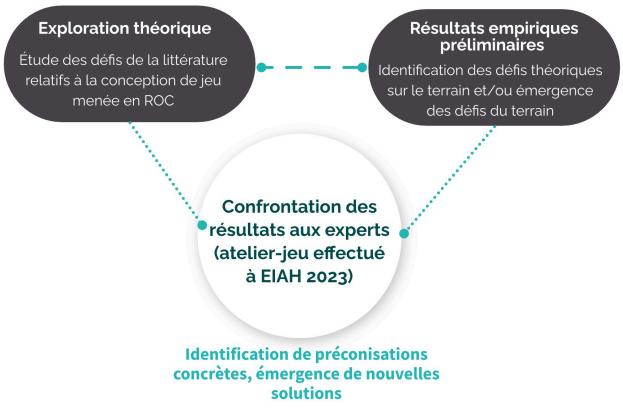


Fig. 1. Défis croisés : de l'exploration théorique aux résultats empiriques

Cet article propose une analyse théorique et empirique des défis, tout en proposant une approche ludique pour les relever. Ainsi, la section 2 articule ces défis identifiés dans la revue de littérature et les résultats préliminaires issus d'un premier terrain d'étude. Puis, la section 3 décrit l'atelier-jeu où ces défis ont été soumis à un panel d'experts et présente les préconisations proposées. Enfin, la section 4 discute les résultats et ouvre sur le partage des savoirs et la co-problématisation.

2 Défis pour mener un projet de recherche orientée par la conception

Les défis que nous présentons proviennent d'une analyse croisée entre une revue de littérature et des observations réalisées dans un projet de conception et d'évaluation du

jeu TSADK destiné aux étudiants de licence 3 pour apprendre la programmation en Java. Le projet réunit 12 acteurs provenant de 3 institutions : 1 chef de projet (P), 1 directeur de filière d'enseignement (H), 3 chercheurs (C1, C2, C3), 2 doctorants (D1, D2), 2 enseignants (E1, E2), et 3 game designers (G1, G2, G3). Ces acteurs sont issus de disciplines différentes telles que la didactique de l'informatique, la recherche en informatique, l'ingénierie pédagogique, la conception de jeux, les méthodologies de recherche, les sciences de l'éducation. Nous avons suivi 2 itérations du jeu TSADK dédiées à la conception et l'évaluation du jeu, au cours de 54 réunions (R#), menées par zoom, enregistrées et retranscrites. La première itération a permis de produire le jeu, la seconde de l'améliorer.

Défi 1. Négocier les intérêts et gérer les attentes divergentes entre le monde de la recherche et celui de la pratique. Les acteurs impliqués dans une ROC, qu'ils soient issus du monde de la recherche ou de la pratique, peuvent avoir des objectifs variés pouvant freiner la collaboration [5]. Ces divergences apparaissent particulièrement au démarrage du projet, amenant les acteurs à expliciter leurs attentes dès le départ sur différents éléments (e.g. l'opérationnalisation des objectifs pédagogiques, la co-problématisation de la question de recherche initiale, ou tout autre élément nécessitant un accord pour collaborer). Un enseignant l'exprime lors d'une réunion : "*J'ai l'impression qu'on n'est pas du tout sur les mêmes longueurs d'onde [...] entre G1, qui elle pense en fonction de ce qu'elle sait faire, moi, en fonction de mes connaissances du jeu, les mecs qui sont en pédagogie qui ont aucune vision de l'informatique, [...] je pense qu'on n'est pas du tout en phase avec tout le groupe.*" (E2, R7). Ce défi porte sur la clarification et l'alignement des attentes pour établir une base commune. Il souligne également que les savoirs professionnels des praticiens sont tout aussi valides et indispensables que les savoirs scientifiques [6].

Défi 2. Construire et maintenir une vision commune du processus. Un des défis de la recherche collaborative est de construire et maintenir une vision commune du processus [7]. Cette vision partagée guide les acteurs en organisant ce processus [8] et en clarifiant les liens entre les éléments [9]. Par exemple, il y a un enjeu à ce que les acteurs partagent une vision commune des objectifs pédagogiques et les comprennent [10]. Cet enjeu a été identifié dans la réunion (R5) :

- C3 : "*J'ai un peu de mal à voir comment on évalue cette chose-là [...] j'arrive pas à me projeter sur un objectif comme ça [...].*"
- P : "*Ce qu'on peut regarder c'est l'objectif 1 et le décortiquer complètement [...].*"
- C2 : "*Je pense qu'aussi pour bien travailler ensemble [...], il faut qu'ils [les non informaticiens] comprennent bien l'objectif, [...] pour pas qu'on fasse des propositions dans la discussion qui soient complètement à côté de la plaque*"

Ce défi est également saillant lorsque chercheurs et praticiens négocient une problématique et des questions de recherche. Cela exige d'équilibrer attentes pratiques et académiques [11], tout en reflétant les apports et besoins des différents acteurs. Ceci implique des ajustements pour conserver une cohérence dans le processus collaboratif.

Défi 3. Articuler des éléments ludiques et pédagogiques. Pour concevoir un jeu éducatif, il est recommandé d'articuler les éléments ludiques et pédagogiques [12]. Si cette articulation n'est pas aboutie, cela peut entraîner une perte de sens entre le contenu à apprendre et ce que le joueur fait dans le jeu [7]. La première itération du jeu TSADK a été pré-testée avec 9 étudiants, aux caractéristiques proches des futurs apprenants-joueurs. Les retours ont été unanimes, ils n'ont pas eu l'impression de jouer : "Y*'en a qui ont dit, c'est pire qu'un TP classique, pour certains, ça les frustrait d'être dans un environnement où au final, c'est quand même un peu plus limitant qu'un éditeur de codes classique [...] mais aucun qui a vraiment eu l'impression de jouer*" (E2, R27). Lors de la seconde itération du jeu, un travail d'amélioration des éléments ludiques a été mené et leur lien avec les éléments pédagogiques a été renforcé : "*Il faudra aligner avec les pré-requis pédagogiques [...] il faudra dire on peut débloquer tel indice [à] tel moment. [...] Le déblocage des salles sera pensé en même temps que les activités pédagogiques.*" (E2, R29).

Défi 4. Élaborer des indicateurs. Définir des indicateurs pertinents permettant d'évaluer si le jeu atteint les objectifs visés est également un défi. Ces indicateurs répondent non seulement aux objectifs pédagogiques des praticiens, mais servent aussi les besoins de la recherche. Ils doivent également refléter les spécificités des jeux. En recherche collaborative, ce processus est complexe car les indicateurs doivent être compréhensibles et exploitables par tous les acteurs. Souvent, leur conception est mal alignée avec les besoins des utilisateurs [13], et repose excessivement sur la quantité de données disponibles dans les environnements d'apprentissage au détriment de leur pertinence [14]. Un équilibre est difficile à atteindre entre les attentes des praticiens, qui souhaitent des indicateurs opérationnels adaptés au contexte éducatif, et celles des chercheurs, qui privilégient des mesures alignées avec leurs modèles théoriques. Dans le projet que nous avons suivi, cette difficulté est apparue dans les discussions sur l'articulation entre les mesures factuelles (traces d'interaction) et les données des questionnaires : "*Ce qui m'intéresserait de savoir, ce seraient quels sont les leviers qui poussent l'étudiant à devenir autonome. Est-ce que c'est le levier ludique [ou] autre chose ?"* (E2, R40). Par ailleurs, le modèle théorique, introduit après la réalisation du premier prototype du jeu, n'a pas été intégré dans le scénario ni dans les énigmes, compliquant l'alignement des choix de conception du jeu avec les concepts théoriques.

Défi 5. Identifier les compétences manquantes. Le manque de certaines compétences peut freiner le processus collaboratif. Même s'il est recommandé d'identifier les compétences requises en amont du processus, il est parfois difficile d'anticiper toutes celles qui le seront, car des besoins spécifiques émergent souvent durant la conception et l'évaluation du jeu. Ainsi, le game designer (G1) a exprimé les limites de ses compétences techniques pour mettre en place un système de traçage : "*J'ai essayé de manipuler un peu WEGAS [plateforme de développement de jeu] et j'arrive à faire des choses basiques. Mais quand je vois les jeux qui sont déjà sur le marché, même des trucs parfois pas très compliqués pour l'instant, je ne suis pas capable de faire ce genre de chose avec WEGAS, [...] il y aurait quelqu'un d'autre qui serait là, un développeur qui ferait*

des gros trucs ? [...] pour les implémenter, j'ai pour l'instant, aucune idée de comment faire ça." (R6). L'enseignant (E2) a également souligné les conséquences de ces limites sur la conception du jeu et ses futures itérations : "*Il va falloir que tu [G1] réfléchisses comment gérer ça parce que en fait des combinaisons t'en auras dans tous les sens, [...] si déjà t'as pas débloqué ces verrous-là, pour moi, le jeu sera jamais évolutif.*" (R13).

Défi 6. Identifier un responsable par phase selon ses expertises. Ce défi souligne l'importance de désigner un responsable pour chaque phase clé du projet. Cela renvoie à la notion de stage leader [15], située entre le broker et le chef de projet. Par sa connaissance suffisante des savoirs détenus par les acteurs qui collaborent, le broker facilite leur compréhension mutuelle [16]. Le chef de projet gère la complexité du projet et ses imprévus [17]. Ainsi, le stage leader soutient la collaboration entre les acteurs travaillant sur sa phase (e.g. veille à sa cohérence avec le projet et à la communication entre eux). Les difficultés liées à son absence se sont illustrées à plusieurs reprises dans TSADK, comme lors de la construction d'une métaphore du jeu (défi 3) abordé en réunion (R5) :

- C2 : "On a l'habitude de travailler sur [...] une espèce de métaphore globale qui représente la situation que l'on veut [enseigner], qui est très difficile à faire, [...] Mais c'est un boulot d'informaticien pour dire : Mais c'est ça qui est important quoi [...]."

- E1 : "Et nous, on n'est pas capable de trouver les bonnes métaphores donc euh on se complète bien."

Aucun des participants n'a pris en charge la conduite de cette phase, et malgré le travail effectué, le jeu TSADK ne contient pas de métaphore au sens de [18]. De ce constat est né le besoin qu'un expert de la construction de la métaphore pilote cette phase.

Certains défis présentés sont particulièrement approfondis dans la littérature (défis 1 et 2). Ces défis ont également été observés lors de l'atelier-jeu à EIAH2023. D'autres ont certes été identifiés dans la littérature mais ils ont été bien plus approfondis et contextualisés sur ce terrain (défis 3 et 4). Enfin, les défis 5 et 6, bien qu'évoqués de manière générale dans certains travaux, ont été davantage spécifiés grâce à ce contexte empirique.

3 Mise en discussion des défis lors d'un atelier-jeu mené à EIAH2023

Pour étudier plus profondément les défis susmentionnés, nous avons conçu et animé un atelier durant la conférence EIAH2023, puis analysé les résultats des échanges.

3.1 Méthode de conduite de la recherche : description de l'atelier-jeu

L'atelier proposé sous forme d'un jeu, a mis en scène les défis décrits précédemment. Il visait à mobiliser l'expertise des participants afin qu'ils formulent des préconisations

pour résoudre ces défis. Ces participants présentaient des profils variés : ingénieurs pédagogiques, chercheurs en informatique, experts en conception de jeu, enseignants, chercheurs en sciences de l'éducation.

Le choix de ce format ludique était motivé par plusieurs raisons : 1) favoriser l'engagement des participants autour d'une expérience commune, 2) encourager une discussion égalitaire où chaque participant dispose du même statut, celui de joueur, 3) soutenir la collaboration durant l'atelier en proposant un jeu nécessitant la résolution des énigmes en équipe, 4) créer un espace de créativité pour résoudre des problèmes complexes, en favorisant une réflexion collective pendant le jeu et lors du débriefing.

Ainsi, notre atelier a été conçu comme une méthode intégrée à notre démarche de recherche pour valider empiriquement les défis identifiés et faire émerger des préconisations. Comme le soulignent [19], aucun expert ne détient à lui seul tout le savoir nécessaire pour résoudre ces défis. La diversité des points de vue révèle des éléments invisibles pour certains mais saillants pour d'autres.

Nous ne pouvons pas affirmer l'atteinte de ces objectifs, toutefois les participants ont salué l'originalité de l'atelier-jeu.

Dans ce jeu, les participants sont enfermés dans la pyramide d'un dieu égyptien. Pour se libérer, ils doivent accomplir une mission : résoudre les difficultés liées à la conception d'un jeu éducatif. Pendant une heure, les participants devaient 1) identifier les rôles, 2) résoudre des énigmes représentant des défis et 3) rédiger leurs préconisations sur un papyrus pour guider les futures équipes.

Les 21 participants, répartis en cinq équipes (1 groupe de 5 participants et 4 groupes de 4 participants), ont travaillé sur trois plateaux thématiques. Les groupes 1 (A1 à A4) et 2 (A5 à A9) ont utilisé le plateau "démarrer la conception de jeu" constitué de 4 énigmes, et le groupe 3 (A10 à A13) a exploré le plateau "opérationnaliser les objectifs pédagogiques et recourir à une métaphore" qui comptait 4 énigmes. Ces deux plateaux visent à étudier les trois phases sur lesquelles se centrent le premier sujet de thèse, portant sur les dynamiques collaboratives lors de la conception de jeu. Les groupes 4 (A14 à A17) et 5 (A18 à A21) ont travaillé sur le plateau "conduire une recherche pour produire des traces d'interaction" composé de 7 énigmes. Ce plateau vise à expérimenter le deuxième sujet de thèse, axé sur la co-spécification des traces numériques d'interaction. Chaque plateau comprenait cartes de jeu, pions, dés, sablier pour limiter le temps, et papyrus pour rédiger les préconisations (voir fig. 2). Un débriefing collectif a permis aux équipes de partager leurs préconisations.



Fig. 2. Photo d'un plateau de jeu.

L'ensemble des échanges, intra et inter-groupe ont été enregistrés, transcrits, puis soumis à une analyse thématique [20]. Le jeu de plateau nous a aussi permis de collecter des données (e.g. les groupes pouvaient annoter les cartes du jeu).

3.2 Résultats : synthèse des préconisations issues de l'atelier-jeu

Les échanges recueillis durant l'atelier-jeu ont été analysés afin d'identifier des préconisations pour relever les défis décrits précédemment. Ces résultats, synthétisés dans la Table 1, reflètent les perspectives interdisciplinaires des participants (A#).

Préconisations du défi 1. Négocier les intérêts et gérer les attentes divergentes. L'une des préconisations évoquées par les experts pour gérer les attentes variées concerne l'utilisation d'outils de communication, tels que ceux issus des méthodes agiles :

- A9 : "*On utilise une méthodologie agile, donc on va utiliser des outils genre Miro, donc les grands board interactifs où ils vont utiliser certaines méthodes comme Kanban.*"
- A7 : "*Donc la solution ce serait d'utiliser des outils de collaboration ?*"
- A9 : "*[...] Ouais [...] des méthodes d'échange, de visualisation, en fait de qui fait quoi."*

Préconisations du défi 2. Construire et maintenir une vision commune du processus. Les discussions reflètent la mise en place d'un lexique commun et l'usage des stratégies de vulgarisation pour réduire les malentendus entre disciplines, comme le souligne le groupe 1 : "*C'est vraiment négocier entre les différentes personnes qui ont des registres de connaissance différents pour asseoir un pot commun*" (A1). En parallèle, le groupe 2 propose de "*se mettre d'accord en fait sur un but à atteindre*" (A9) et "*[d'aligner les]objectifs pour tout le monde*" (A7).

Préconisations du défi 3. Articuler des éléments ludiques et pédagogiques. Ces préconisations concernent par exemple la définition de ce qui a trait au ludique, comme

l'évoque le participant A12 (groupe 3) : "*Le pédagogique c'est très bien défini, pour l'instant rien n'est défini sur ce qu'est le ludique [...]. Pour moi y a deux dimensions possibles, [...] il peut y avoir de la gamification, c'est-à-dire rajouter sur quelque chose qui est pas du tout ludique, tout un tas de stratégies pour le ludifier ou alors penser dès le départ une articulation entre les objectifs pédagogiques [...] et le ludique."*

Préconisations du défi 4. Élaborer des indicateurs. L'élaboration des indicateurs adaptés a suscité des discussions sur les termes clés "mesure", "variable" et "indicateur", reflétant la diversité des perspectives disciplinaires des participants. Pour éviter les malentendus terminologiques, le participant A17 (groupe 4) propose : "*[...] Oui, du cadre épistémologique. Si on ne s'est pas mis d'accord sur le cadre théorique, ça ne va pas bien."*

Préconisations du défi 5. Identifier les compétences manquantes. Pour relever ce défi, le groupe 1 propose par exemple de former une équipe disposant des compétences nécessaires : "*La solution c'est de créer une équipe projet pluridisciplinaire avec toutes les compétences dont on a besoin"* (A3).

Préconisations du défi 6. Identifier un responsable par phase selon ses expertises. Les discussions ont tourné autour de l'importance de désigner un responsable pour coordonner les phases du projet. Les verbatims mentionnent parfois le chef de projet, d'autres un responsable de recherche, quelqu'un qui doit comprendre les spécificités des disciplines impliquées. Cependant, les préconisations ne s'alignent pas directement sur la définition conceptuelle proposée pour ce défi.

Table 1. Préconisations issues de l'atelier-jeu.

Défis	Préconisations
Défi 1. Gestion des attentes divergentes	- Planifier une phase pour exprimer les attentes respectives - Proposer des outils pour gérer la communication - Désigner un responsable pour piloter la conduite de la recherche
Défi 2. Vision commune du processus	- Définir les termes et concepts clés des disciplines - Se mettre d'accord sur le but à atteindre - Prioriser les objectifs pédagogiques selon leur difficulté - Introduire un rôle de médiation des techniques d'animation
Défi 3. Articulation ludo-pédagogiques	- Définir ce qui a trait au ludique (e.g identifier les mécaniques ludiques)

	- Construire le débriefing du jeu pour qu'il y ait apprentissage - Expliquer le concept de métaphore car difficile à appréhender
Défi 4. Élaboration des indicateurs	- Établir un cadre théorique clair et partagé - Définir un protocole d'expérimentation claire et des méthodes de collecte de données
Défi 5. Compétences manquantes	- Former une équipe disposant des compétences nécessaires (game design, développement, UX, analyse de données, savoirs disciplinaires, pédagogie) - Réaliser une analyse des compétences existantes pour combler les lacunes
Défi 6. Responsable par phase	- Désigner un responsable pour coordonner les phases du projet (e.g chef de projet, responsable de recherche) - Comprendre les spécificités des disciplines impliquées

Une limite se pose à l'analyse des préconisations. Certaines énigmes, jugées trop longues, ont limité le temps de discussion, entraînant parfois des échanges superficiels.

4 Discussion et perspectives

Cet article explore les défis rencontrés par les équipes pluridisciplinaires dans la conduite collaborative d'une ROC menée pour concevoir les jeux éducatifs. Une revue de littérature et un terrain TSADK ont permis d'identifier plusieurs types de défis. Certains concernent la conduite de projet en général, d'autres sont spécifiques à la conception ou à l'évaluation d'un jeu, alors que certains relèvent de la recherche elle-même. Plusieurs défis n'ont pas été discutés ici, comme ceux relatifs aux principes éthiques pour concevoir un dispositif d'apprentissage respectueux des droits des personnes (e.g. protection des données personnelles des apprenants).

Nous avons discuté six défis majeurs, dont quatre explicitement décrits dans la littérature (défis 1 à 4), et observés sur le terrain TSADK. En revanche, les défis liés aux compétences manquantes et à la gestion des expertises (défis 5 et 6) émergent principalement du contexte empirique.

Ces défis ont ensuite été confrontés aux experts lors d'un atelier-jeu effectué à la conférence EIAH2023, qui ont reconnu les avoir rencontrés dans leur pratique, confirmant leur pertinence théorique et empirique. Pour relever ces défis, ils ont discuté et formulé des préconisations pour soutenir la conception de jeux en ROC et pourraient être portées à la connaissance des équipes pluridisciplinaires dès le début des projets.

Toutefois, certaines préconisations demeurent générales ou incomplètes et nécessitent un approfondissement pour relever les défis identifiés. Par exemple, la négociation

des intérêts variés (défi 1) et la construction d'une vision partagée (défi 2), peuvent être soutenus par des outils et méthodes collaboratifs. Mais sans cadre de confiance, les outils restent insuffisants [21]. De même, peu de solutions concrètes ont émergé et des confusions subsistent autour de l'élaboration des indicateurs (défi 4), entre leur aspect mesurable et construit. Des termes comme "mesure", "valeur", "indicateur", n'ont pas fait consensus reflétant ainsi les différences disciplinaires. Cette divergence souligne la difficulté de construire un langage commun en ROC.

Par ailleurs, l'absence de préconisations explicites sur les compétences manquantes (défi 5) et la conduite des phases (défi 6), nous amène à souligner l'importance du partage des savoirs au sein de la ROC. Le savoir permet de porter un éclairage nouveau sur son environnement et d'évoquer ses connaissances à travers un langage adéquat [22]. Pour que les acteurs puissent se comprendre pour collaborer lors de la conception de jeux, ils sont amenés à partager leurs savoirs [23]. Ces savoirs sont transformés pour être mobilisés [24]. Ce partage est difficile car les acteurs peuvent utiliser des terminologies ou significations différentes pouvant freiner leur compréhension mutuelle [25]; ils n'ont pas toujours conscience des savoirs qu'ils détiennent [26], complexifiant leur explicitation et leur mobilisation ; enfin, un décalage peut émerger entre l'explicitation d'un savoir théorique (e.g. métaphore d'un jeu) et sa mise en pratique (e.g. construction d'une métaphore d'un jeu). Outre le lien avec les défis 5 et 6, cela nous questionne plus largement sur les collaborations pluridisciplinaires : *Comment lever les blocages liés au partage des savoirs ? Comment formaliser ce partage pour soutenir les équipes dans la conception et l'évaluation des jeux éducatifs ? En quoi peut-il devenir un levier concret ?*

En parallèle, l'analyse des préconisations révèle également des limites liées à la problématisation d'une question éducative. Ce processus de co-problématisation, bien que fondamental pour effectuer des choix méthodologiques [27] et structurer les pratiques collaboratives [28], est encore peu formalisé dans les travaux sur la conception de jeux éducatifs et la ROC [29]. En lien avec ce processus, l'élaboration des indicateurs reste également un défi comme expliqué précédemment. Ces indicateurs doivent refléter à la fois les objectifs pédagogiques des praticiens et les objectifs de recherche des chercheurs, en considérant les spécificités des jeux éducatifs. Cette étape de la co-problématisation est souvent fragmentée, ce qui peut freiner l'alignement des indicateurs avec les besoins des praticiens et les modèles théoriques des chercheurs. Ces observations soulèvent deux questions : *Comment définir et modéliser la "co-problématisation" d'une question éducative ? Comment structurer et formaliser la co-élaboration des indicateurs afin qu'ils deviennent adaptés aux pratiques pédagogiques et ancrés dans des modèles théoriques ?*

Enfin, la structuration de l'atelier en plateaux distincts a limité la prise en compte des interactions entre les défis et de la dimension systémique de la ROC. Cette séparation, conçue pour alimenter deux thèses distinctes, a fragmenté la réflexion des experts et limité une analyse globale des défis.

En conclusion, cet article explore les défis rencontrés par les équipes pluridisciplinaires dans la conduite d'une ROC menée pour concevoir les jeux éducatifs. Les préconisations formulées par les experts valident la pertinence des défis identifiés, et peuvent soutenir les acteurs débutants en conception de jeu. Cependant, elles s'avèrent trop

générales ou insuffisamment adaptables aux besoins spécifiques des acteurs expérimentés en conception de jeux menée en ROC. Cette observation souligne la nécessité d'approfondir certains défis, notamment ceux liés à la gestion des expertises, au partage des savoirs, à la formalisation des questions de recherche (co-problématisation) et à la co-élaboration des indicateurs. Dans cette perspective, nous poursuivons l'étude de deux processus clés : 1) la co-problématisation et l'élaboration des indicateurs et 2) le partage de savoirs entre les acteurs de la conception. Ces démarches, encore peu formalisées et outillées dans les travaux actuels de la ROC, représentent un levier pour renforcer la cohérence entre rigueur scientifique de la démarche de recherche et conception des jeux éducatifs.

Remerciement. Nous remercions l'équipe du projet TSADK, les participants à l'atelier-jeu de la conférence EIAH2023, et le projet co.LAB dans lequel s'inscrivent ces deux terrains d'étude. Ce projet a été financé par le Fonds National Suisse (PNR77).

Ethique. Ce projet a reçu l'approbation de la commission éthique de l'Université de Genève (CUREG-2021-02-18).

References

1. Sanchez, E., Monod-Ansaldi, R. : Recherche collaborative orientée par la conception. Education & didactique 9(2), 73-94 (2015).
2. Sensevy, G.: Le sens du savoir. De Boeck, Bruxelles (2011).
3. Morrissette, J., Compaoré, G.: Le savoir faire enseignant sur l'évaluation formative informelle. Formation et profession 20(3), 54–64 (2013).
4. Sanchez, É., Prior, E., Mandran, N., Monnier, S., Huguenin, A., Laurent, M., Jaouadi, M., Jaccard, D.: Co.LAB, une plateforme d'accompagnement pour la conception et l'évaluation collaboratives de jeux destinés à un usage éducatif. In: Actes de la onzième Conférence sur les Environnements Informatiques pour l'Apprentissage Humain (EIAH), pp. 80–85. Brest, France (2023).
5. Klerks, G., Slingerland, G., Kalinauskaite, I.: When Reality Kicks In: Exploring the Influence of Local Context on Community-Based Design. Sustainability 14(7), 4107 (2022).
6. Elisme Pierre, E., Girardet, C., Mottier Lopez, L.: Défis de la co-situation d'une recherche collaborative: une expérience de co-élaboration, en partie à distance, d'un dispositif d'évaluation formative pour des élèves allophones en Guyane française. La Revue LEeE 3, 1–30 (2021).
7. Ke, F., Shute, V., Clark, K. M., Erlebacher, G.: Interdisciplinary Design of Game-Based Learning Platforms: A Phenomenological Examination of the Integrative Design of Game, Learning, and Assessment. Springer, Cham (2019).
8. Laakso, N. L., Korhonen, T. S., Hakkarainen, K. P. J.: Developing students' digital competences through collaborative game design. Computers & Education 174, 104308 (2021).
9. Carvalho, M.B., Bellotti, F., Berta, R., De Gloria, A., Sedano, C.I., Hauge, J.B., Hu, J., Rauterberg, M.: An activity theory-based model for serious games analysis and conceptual design. Computers & Education 87, 166–181 (2015).
10. Lynch, M. A., Tunstall, R. J.: When worlds collide: Developing game-design partnerships in universities. Simulation & Gaming 39(3), 379–398 (2008).

11. Barry, S., Saboya, M., Corriveau, C., Bednarz, N., Maheux, J.-F.: Défis et enjeux de la démarche de recherche collaborative en didactique des mathématiques. In: Bronner, A., Bulf, C., Castela, C., Georget, J.P., Larguier, M., Pedemonte, B., Pressiat, A., Roditi, É. (eds.) Questions vives en didactique des mathématiques: problèmes de la profession d'enseignant, rôle du langage. 16e école d'été de didactique des mathématiques, pp. 447–482. La Pensée Sauvage, Grenoble (2012).
12. Arnab, S., Lim, T., Carvalho, M.B., Bellotti, F., Freitas, S. de, Louchart, S., Suttie, N., Berta, R., Gloria, A.D.: Mapping learning and game mechanics for serious games analysis. British Journal of Educational Technology 46(2), 391–411 (2015).
13. Chatti, M.A., Muslim, A., Guesmi, M., Richtscheid, F., Nasimi, D., Shahin, A., Damera, R.: How to Design Effective Learning Analytics Indicators? A Human-Centered Design Approach. In: Alario-Hoyos, C., Rodríguez-Triana, M.J., Scheffel, M., Arnedillo-Sánchez, I., Dennerlein, S.M. (eds.) Addressing Global Challenges and Quality Education. EC-TEL 2020, Lecture Notes in Computer Science, vol. 12315. Springer, Cham (2020).
14. Dyckhoff, A.L.: Action Research and Learning Analytics in Higher Education. RWTH Aachen University, (2014).
15. Mandran, N., Prior, E., Sanchez, E.,Vermeulen, M.: Reorienting Learning Game Design in Design-Based Research: a Case Study. arXiv (2024).
16. Monod-Ansaldi, R., Vincent, C., Aldon, G.: Objets frontières et brokering dans les négociations en recherche orientée par la conception. Éducation et Didactique 13(2), 61–84 (2019).
17. Krahn, J., Hartman, F.: Effective project leadership: a combination of project manager skills and competencies in context. In: PMI® Research Conference: New Directions in Project Management. Project Management Institute, Newtown Square, PA (2006).
18. Bonnat, C., Sanchez, E.: Toward a Digital Companion to Monitor a Mixed Reality Game. International Journal of Serious Games 9(3), 5–21 (2022).
19. Abrahamson, D., Chase, K.: Interfacing practices: domain theory emerges via collaborative reflection. Reflective Practice 16(3), 372–389 (2015).
20. Paillé, P., Mucchielli, A.: Chapitre 11. L'analyse thématique. In: L'analyse qualitative en sciences humaines et sociales, pp. 235–312. Armand Colin (2016).
21. Alenljung, B., Maurin Söderholm, H.: Designing Simulation-Based Training for Prehospital Emergency Care: Participation from a Participant Perspective. In: Kurosu, M. (ed.) Human-Computer Interaction: Design and Evaluation. HCI 2015, Lecture Notes in Computer Science, vol. 9169, pp. 297–306. Springer, Cham (2015).
22. Astolfi, J.-P.: La saveur des savoirs: disciplines et plaisir d'apprendre. 5e edn. ESF Sciences humaines, Paris (2021).
23. Monod-Ansaldi, R., Vincent, C., Aldon, G.: Objets frontières et brokering dans les négociations en recherche orientée par la conception. Education et didactique 13(2), 61–84 (2019).
24. Ligozat, F., Charmillot, M., Muller, A.: Acteurs de la recherche en éducation: quels partages des savoirs? In: Le partage des savoirs dans les processus de recherche en éducation, pp. 7–32. De Boeck Supérieur, Louvain-la-Neuve (2016).
25. Carlile, P. R.: Transferring, Translating, and Transforming: An Integrative Framework for Managing Knowledge Across Boundaries. Organization Science 15(5), 555–568 (2004).
26. Roungas, B., Bekius, F., Verbraeck, A., Meijer, S.: Improving the decisionmaking qualities of gaming simulations. Journal of Simulation 15(3), 177–190 (2021).
27. Kivits, J., Balard, F.: Chapitre 2. La problématisation ou l'importance de penser sa question de recherche. In: Les recherches qualitatives en santé, pp. 43–59. Armand Colin, (2016).
28. Jaouadi, M.: Co-problématiser dans le cadre d'une recherche orientée par la conception. In: Actes des neuvièmes rencontres jeunes chercheur.e.s en EIAH, pp. 76–83. Lille, France (2022).

29. Jaouadi, M., Bonvin, G., Parmentier, H., Sénécail, A.: Design de la recherche: Problématiser, expérimenter, produire et interpréter les données. ISTE (en préparation 2025).

Modélisation et décision de rétroactions épistémiques dans l'environnement AlgoPython

Sébastien Jolivet[1,2][0000−0003−3915−8465], Badmavasan Kirouchenassamy[2][0009−0003−6502−154X], Amel Yessad[2][0000−0001−7575−6433], and Vanda Luengo[2][0000−0003−3915−8465]

[1] IUFE & TECFA, , Université de Genève, Suisse
sebastien.jolivet@unige.ch
[2] Sorbonne Université, CNRS, LIP6, F-75005 Paris, France
{prenom.nom}@lip6.fr

Résumé AlgoPython est une plateforme d'apprentissage de la programmation en Python. Dans sa version actuelle elle propose des rétroactions limitées aux erreurs syntaxiques et à la validation de la tâche. Dans cet article nous étudions les différents éléments nécessaires à la mise en place d'un système de décision de rétroactions épistémiques adaptées à l'apprenant. Après avoir présenté le problème de la décision des rétroactions, nous analysons un algorithme de type *Reinforcement learning* pour l'aborder. Nous définissons alors deux éléments fondamentaux : la définition des éléments à prendre en compte pour permettre la décision (éléments relatifs à l'apprenant et à sa production) et la modélisation et la production des rétroactions. Enfin nous proposons une première preuve de concept permettant d'évaluer la production des rétroactions ainsi que le choix d'un algorithme de type *Reinforcement learning* pour la décision de rétroactions.

Keywords: apprentissage de la programmation · rétroactions épistémiques · décision · apprentissage par renforcement

Abstract. AlgoPython is a platform designed to teach Python programming. Currently, it only provides feedback on syntax errors and task validation. This article studies the elements necessary for implementing a system that provides adaptive epistemic feedback. After presenting the feedback decision problem, we analyze a reinforcement learning algorithm to address it. Next, we define two fundamental elements: the elements to be considered for decision (elements relating to the learner and their output) and feedback modeling and production. Finally, we propose an initial proof of concept to evaluate feedback production and reinforcement learning algorithm selection for feedback decision.

Keywords: computer science education · programming learning · epistemic feedback · decision · reinforcement learning

1 Introduction

AlgoPython [3] est un environnement d'apprentissage de la programmation en Python qui, dans sa version actuelle, propose des rétroactions limitées. Elles portent essentiellement sur la syntaxe ou la (non)validation de la réponse, et ne sont pas adaptées à l'apprenant. Or, la littérature met en évidence deux éléments ; d'une part les rétroactions les plus efficaces sont celles liées à la tâche [30] et d'autre part il y a un impact positif de l'adaptation des EIAH aux caractéristiques des apprenants [27,29,1].

Dans le cadre de cet article, nous nous intéressons à la mise en place d'un système permettant de proposer des rétroactions relatives aux tâches et adaptées aux apprenants dans AlgoPython.

Pour illustrer la complexité du problème considérons l'exemple suivant. Un apprenant doit réaliser l'exercice proposé Fig. 1. Comme réponse il soumet le code présenté Fig. 2.

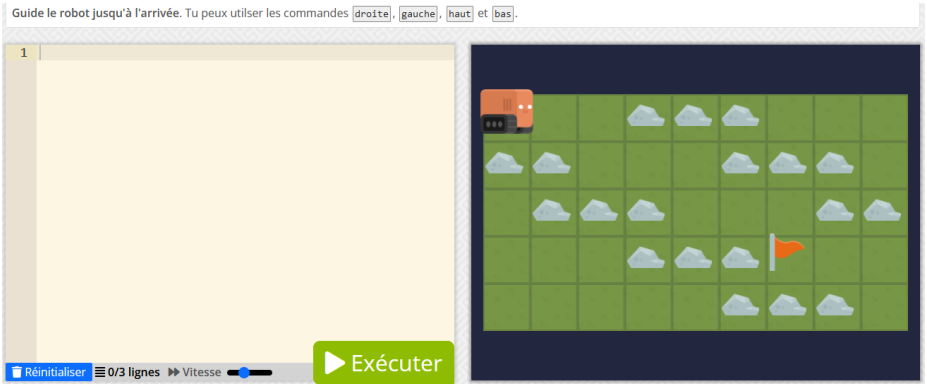


FIGURE 1 : Exercice de déplacement du robot

Même pour un exercice aussi simple, et pour une réponse aussi courte, les critères pouvant guider le contenu et le choix de la rétroaction sont nombreux : accompagner la modélisation de la situation ; aider tout d'abord à corriger l'erreur relative au compteur de la boucle ou au corps de la boucle ; proposer un rappel de cours ou proposer un exemple proche de la situation à traiter ; signaler explicitement à l'apprenant une erreur ou le laisser interpréter l'aide qu'on lui propose ; etc. Par exemple, les trois rétroactions proposées Fig. 8, Fig. 9 et Fig. 10 ont toutes une pertinence potentielle par rapport à la réponse de l'apprenant.

S'ajoute aussi l'observation de Nogry & al. "Dans le cadre des EIAH, l'objectif à atteindre comporte deux niveaux : l'apprentissage (de la discipline enseignée et non de la manipulation du système) et la réalisation de tâches proposées par le système (résolution de problème, recherche d'informations, simulations...).

[3] AlgoPython : https ://www.algopython.fr, consulté le 3 mai 2025



FIGURE 2 : Par rapport à l'exercice présenté Fig. 1 : solution proposée par un apprenant, à gauche, et solution attendue, à droite.

Même si ces niveaux sont connectés, il n'y a pas de lien direct entre la réalisation de la tâche et l'apprentissage effectif, un échec dans la réalisation de la tâche peut, dans certaines conditions, être bénéfique pour l'apprentissage" [19] (p. 267).

La complexité mise en évidence par cet exemple peut se résumer ainsi : *quelle est la bonne rétroaction à proposer à un apprenant, à un moment donné de son apprentissage, par rapport à la réponse qu'il a proposée pour permettre la réalisation de la tâche et l'apprentissage ?*

L'objectif global de notre travail est d'apporter une réponse à cette question en implémentant un système de décision de rétroactions épistémiques et adaptées, soit des rétroactions centrées sur la tâche à réaliser, les connaissances en jeu et qui prennent en compte l'apprenant et son activité. Nous présentons de premiers éléments dans cette contribution.

2 Questions de recherche

Il existe une littérature abondante relative aux rétroactions. On pourra par exemple se référer à [11] ou [30] pour les rétroactions en éducation en général, [23] propose un focus particulier sur les rétroactions formatives et [15] propose une méta-analyse des effets des rétroactions dans un environnement numérique. [24] et [14] proposent des éléments spécifiques aux rétroactions dans le contexte de l'apprentissage de la programmation.

La question identifiée à la fin de l'introduction peut être abordée selon différents angles. Au regard des données déjà disponibles, et des traces accessibles depuis la plateforme, nous avons fait le choix de nous intéresser particulièrement aux productions de l'apprenant (plusieurs milliers de soumissions de code à différents exercices déjà disponibles au commencement de notre travail) et aux informations que l'on peut en déduire. Dans cet article nous nous concentrons sur la question des erreurs. Elles vont être exploitées à la fois comme critère de décision de la rétroaction et participer à caractériser le profil de l'apprenant.

Les différentes étapes du processus de décision sont décrites dans la figure Fig. 3. L'étape 3 n'est pas traitée dans cet article, nous formalisons les étapes 1 et 2 au travers des questions de recherche suivantes.

(QR1) : quelle modélisation des rétroactions pour permettre leur production et leur décision ?

(QR2) : quel type de modèles de décision pour proposer des rétroactions épistémiques adaptées à l'apprenant ?

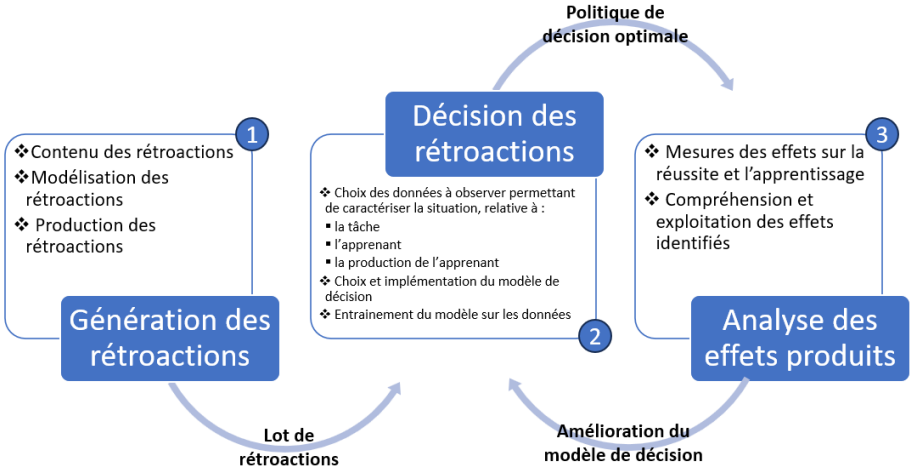


FIGURE 3 : Étapes pour la mise en œuvre d'un système de décision des rétroactions

La seconde est travaillée dans les sections 4 et 5. Les réponses apportées permettent d'aborder la première question dans la section 6.

3 Cadre théorique

Répondre aux questions de recherches 1 et 2 amène à représenter le savoir en jeu, et sa mobilisation, dans les tâches proposées dans AlgoPython. Pour disposer d'un cadre permettant d'aborder de manière commune ces différentes questions nous avons fait le choix d'exploiter la modélisation de l'activité de programmation [4] telle que proposée dans [13]. Appuyée sur le modèle praxéologique de la théorie anthropologique du didactique [7], cette modélisation met en évidence des types de tâches, la manière de les réaliser (nommées *techniques*) et les éléments de savoir (nommés *logos*) utiles pour justifier et guider la mise en œuvre des *techniques*. Ainsi, lorsque l'apprenant réalise une tâche dans AlgoPython, cette tâche appartient à un type de tâches, elle est réalisée par le moyen d'une technique qui mobilise différents éléments du logos. Par exemple lorsque l'apprenant doit réaliser la tâche *Guide le robot jusqu'à l'arrivée* dans le contexte défini (voir Fig. 1) il met en œuvre, notamment, le type de tâches *Concevoir une boucle bornée* qui va être réalisé à l'aide d'une technique pouvant être (partiellement) décrite de la manière suivante : *) identifier la ou les instructions à répéter **) déterminer le nombre de répétition ***) implémenter en Python. Les éléments du logos associés sont la définition d'une boucle bornée, la notion de corps d'une boucle, la notion d'itération et d'itérateur, etc. La figure Fig. 4 illustre cette approche pour deux types de tâches, l'un relatif aux boucles bornées et l'autre relatif aux algorithmes.

[4] Le référentiel est disponible à l'adresse https ://link.infini.fr/ref_prog

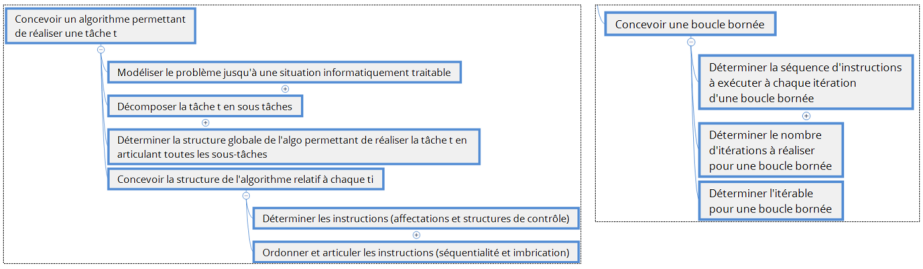


FIGURE 4 : Types de tâches *Concevoir un algorithme permettant de réaliser une tâche t* et *Concevoir une boucle bornée* ainsi que des ingrédients de leurs techniques respectives. Extrait de https ://link.infini.fr/ref_prog

Le référentiel exploité permet de décrire l'activité attendue de l'apprenant mais pas les erreurs de celui-ci. Or, divers travaux relatifs aux rétroactions montrent la pertinence qu'il y a à prendre en compte l'erreur [5,25].

Relativement au processus décrit figure Fig. 3 cette prise en compte peut se situer à la fois dans le contenu des rétroactions, dans la connaissance de l'apprenant et dans la production de l'apprenant. C'est donc un élément que nous considérons comme important pour notre travail.

Par ailleurs, dans la plupart des travaux disponibles relatifs aux erreurs dans l'apprentissage de la programmation il n'y a que des classifications d'assez haut niveau [21]. Or, s'agissant de premiers apprentissages de la programmation, une caractérisation des erreurs trop globale (erreur sur la boucle par exemple) n'est pas efficiente pour orienter le choix de la rétroaction.

Pour conserver le cadre homogène que nous visons, nous exploitons le référentiel cité précédemment pour définir un ensemble d'erreurs. Pour cela, nous associons aux différents types de tâches des erreurs possibles. Par exemple, relativement aux types de tâches présentés Fig. 4, nous définissons les erreurs suivantes : par rapport aux boucles bornées (erreur relative au corps de la boucle ; erreur relative au nombre d'itérations ; erreur dans la définition de l'itérable) ; par rapport aux algorithmes (erreur dans la décomposition du problème en sous-tâches ; erreur dans le choix des structures de contrôle ; erreur dans l'ordre des instructions ; instruction manquante ; instruction inutile ; etc.).

4 Le problème de la décision des rétroactions

Le problème de la décision des rétroactions consiste à choisir, parmi un ensemble de rétroactions disponibles, la plus adaptée à l'apprenant. C'est-à-dire celle qui permet d'atteindre efficacement le ou les objectifs du dispositif (réussite de l'apprenant, meilleur apprentissage des concepts, réduction du temps d'apprentissage, réduction des abondons, etc.).

Or, ce problème est difficile puisque les données (l'état de l'apprenant, les erreurs produites, etc.) sur lesquelles s'appuie la décision sont souvent incertaines,

bruitées et partielles. De plus, l'impact de la décision est, lui aussi, incertain. Ceci nous amène à considérer des modèles d'apprentissage automatique (AA) pour rechercher la meilleure association possible entre une situation (apprenant, production, tâche) et une rétroaction.

Ces modèles AA ont été abondamment utilisés dans le domaine des EIAH (en particulier dans les travaux des communautés AIED et EDM) pour la modélisation de l'apprenant [8,31,17], la prédiction de son comportement [16], l'adaptation des parcours d'apprentissage [1,3], les tests adaptatifs [28], etc. Il existe trois types de modèles d'AA : l'apprentissage non supervisé [10], l'apprentissage supervisé [9] et l'apprentissage par renforcement (Reinforcement Learning ou RL) [26].

Or, les éléments suivants suggèrent une bonne adéquation entre notre problème de décision de rétroactions et les caractéristiques d'un problème RL et :

- Pas de supervision, uniquement un signal de retour/récompense. Contrairement aux modèles d'AA supervisé, ici le modèle n'apprend pas sur des exemples annotés par un superviseur humain puisque les experts ont du mal à choisir la rétroaction la mieux adaptée à la situation de l'apprenant,
- Le retour est généralement retardé et non instantané. En effet, la rétroaction décidée par le système ne doit pas donner nécessairement lieu à une récompense immédiate puisque son effet ne peut être évalué positivement simplement du fait que l'apprenant ait réussi la tâche en cours, des effets sur les apprentissages à moyen et long terme peuvent être privilégiés,
- La prise en compte du temps (les expériences de l'apprenant sont ordonnées dans le temps, les variables ne sont pas indépendantes et identiquement distribuées, non i.i.d),
- Les actions de l'agent RL à l'instant t impactent les données reçues à l'instant $t + 1$. En effet, une rétroaction décidée par le système peut influencer l'apprenant et ses futures tentatives de résolution. Ceci est une caractéristique intrinsèque des modèles RL qui les différencient des modèles supervisés ou non supervisés,
- L'état de l'environnement perçu par l'agent RL est incertain ; Les informations observées sur l'apprenant sont généralement bruitées, incomplètes et incertaines.

Ceci justifie d'écarter la piste des modèles supervisés ou non supervisés et notre choix d'explorer l'approche RL pour apprendre un modèle de décision des rétroactions. Nous présentons maintenant plus précisément cette approche.

Les problèmes RL sont généralement modélisés comme des processus de décision markovien (en anglais Markov decision process, MDP). Un MDP est un quadruplet $\{S, A, T, R\}$ définissant :

- un ensemble d'états S qui peut être fini, dénombrable ou continu ; cet ensemble définit l'environnement tel que perçu par l'agent RL ;
- un ensemble d'actions A qui peut être fini, dénombrable ou continu. L'agent RL interagit avec son environnement en décidant une action de cet ensemble. Dans notre cas, cet ensemble est le corpus des rétroactions disponibles ;

- une fonction de transition $T : S \times A \times S \rightarrow [0; 1]$; cette fonction définit l'effet des actions de l'agent RL sur l'environnement : $T(s, a, s')$ représente la probabilité de se retrouver dans l'état s' en effectuant l'action a, sachant que l'on était précédemment dans l'état s ;
- une fonction de récompense $R : S \times A \times S \rightarrow \mathbb{R}$; elle définit la récompense (positive ou négative) reçue par l'agent pour être passé de l'état s à s' en ayant effectué l'action a. Dans notre cas, la valeur de la récompense dépend de l'objectif de la politique de décision des rétroactions (réussite immédiate de l'apprenant, apprentissage des concepts de programmation, etc.).

Dans la suite nous présentons la construction de la partie $\{S, A\}$ du MDP. La partie $\{T, R\}$ n'est pas abordée dans cette contribution.

5 Définition de l'état de l'environnement du MDP

Un élément clé pour résoudre un problème RL est la définition de l'état de l'environnement. Nous avons choisi de le définir à l'aide du profil des erreurs de l'apprenant, de l'analyse de son code erroné et des caractéristiques de la tâche en cours. Nous ne revenons pas sur cette dernière, elle est basée sur le référentiel présenté section 3. Dans cette même section 3, nous avons défini un ensemble d'erreurs. L'enjeu est de relier un code erroné à une ou plusieurs de ces erreurs. La section suivante détaille les éléments mis en place pour y parvenir.

5.1 Identification des erreurs

Notre approche, pour identifier les erreurs dans un code soumis, est d'exploiter la représentation des codes sous forme d'un arbre syntaxique abstrait (AST) et l'algorithme de Zang et Shasha [32] (AZS). Nous illustrons le travail à l'aide des codes proposés Fig. 2.

Du code à la comparaison d'AST Dans sa version initiale AZS [5] permet de récupérer l'ensemble des opérations de modifications (organisées en trois opérateurs d'édition : insertion ; suppression ; mise à jour) nécessaires pour transformer un AST (celui de l'apprenant) en un autre (un code correct). La difficulté est d'interpréter ces modifications en terme d'erreurs de notre ensemble prédéfini d'erreurs. Nous revenons brièvement sur son traitement.

La bibliothèque Zhang-Shasha par défaut fournit des informations insuffisantes par rapport à nos besoins. Par exemple, si deux instructions conditionnelles se trouvent à des emplacements distincts dans le code, et qu'une erreur est présente dans l'une d'elles, on ne sait pas de laquelle il s'agit dans le code incorrect.

Pour répondre à cette limitation, une méthode a été développée pour extraire et structurer les informations de l'AST sous forme de nœuds adaptés (de l'arbre

[5] https ://github.com/timtadh/zhang-shasha/

AST). Le résultat de ce travail est illustré dans les Fig. 5 et Fig. 6. De plus, l'algorithme de calcul de distance de Zhang-Shasha a été réimplémenté [6] afin d'intégrer le contexte dans lequel les opérations de transformation doivent être appliquées.

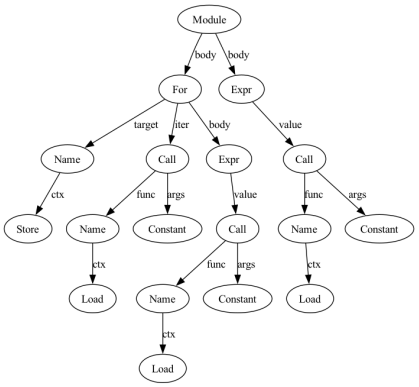


FIGURE 5 : AST sous sa forme originale

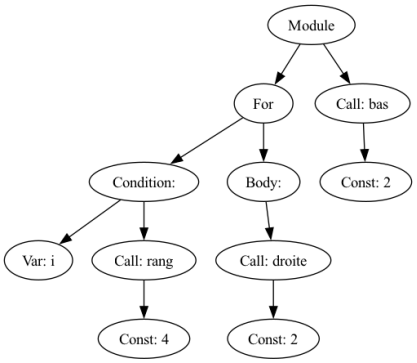


FIGURE 6 : AST reimplementé

En comparant le code incorrect au code correct à l'aide du calcul de la distance de Zhang-Shasha réimplémentée, il est possible de déterminer l'ensemble des opérations nécessaires pour transformer le code incorrect en la solution correcte. Voici ce qui est généré par la comparaison des deux codes de la Fig. 2 :

Listing 1.1 : Liste des opérations d'édition à l'issue de notre réimplémentation de Zang-Shasha dans le cas des deux codes de la Fig. 2

```
{'type': 'update', 'path': ['Module', 'For[0]', 'Condition:[0]', 'Var: i[0]'], 'current': 'Var: i', 'new': 'Var: k'},
{'type': 'update', 'path': ['Module', 'For[0]', 'Condition:[0]', 'Call: range[1]', 'Const: 4[0]'], 'current': 'Const: 4', 'new': 'Const: 3'},
{'type': 'match', 'path': ['Module', 'For[0]', 'Condition:[0]', 'Call: range[1]'], 'current': 'Call: range', 'new': 'Call: range'},
{'type': 'match', 'path': ['Module', 'For[0]', 'Condition:[0]'], 'current': 'Condition:', 'new': 'Condition:'},
{'type': 'match', 'path': ['Module', 'For[0]', 'Body:[1]', 'Call: droite[0]', 'Const: 2[0]'], 'current': 'Const: 2', 'new': 'Const: 2'},
{'type': 'match', 'path': ['Module', 'For[0]', 'Body:[1]', 'Call: droite[0]'], 'current': 'Call: droite', 'new': 'Call: droite'},
{'type': 'insert', 'path': ['Module', 'For[0]', 'Body:[1]', 'Call: bas[1]', 'Const: 1[0]'], 'current': None, 'new': 'Const: 1'},
{'type': 'insert', 'path': ['Module', 'For[0]', 'Body:[1]', 'Call: bas[1]'], 'current': None, 'new': 'Call: bas'},
{'type': 'match', 'path': ['Module', 'For[0]', 'Body:[1]'], 'current': 'Body:', 'new': 'Body:'},
{'type': 'match', 'path': ['Module', 'For[0]'], 'current': 'For', 'new': 'For'},
{'type': 'delete', 'path': ['Module', 'Call: bas[1]', 'Const: 2[0]'], 'current': 'Const: 2', 'new': None},
{'type': 'delete', 'path': ['Module', 'Call: bas[1]'], 'current': 'Call: bas', 'new': None},
{'type': 'match', 'path': ['Module'], 'current': 'Module', 'new': 'Module'}
```

Des AST comparés à l'identification des erreurs Les informations (Voir Listing 1.1) obtenues ne sont pas directement interprétables du point de vue

[6] Ces éléments sont mis à disposition sur https ://pypi.org/project/asterrdetection/

des erreurs. Par exemple, une opération de suppression d'un nœud suivie d'une insertion d'un autre nœud peut indiquer que la section de code concernée n'est pas située au bon endroit. Un autre exemple est le cas où le code apprenant et le code correct diffèrent notamment dans le nommage d'une variable, ceci ne doit pas être interprété comme une erreur alors que des modifications remontent lors de la comparaison des AST.

Pour rendre ces opérations exploitables en tant qu'erreurs logiques de programmation nous avons conçu une méthode permettant de transformer cette liste brute d'opérations d'édition en une liste d'erreurs interprétables, en appliquant un ensemble de règles que nous ne développons pas ici. Le package Python développé est rendu public sur pypi.org et est disponible ici [7].

Voici le résultat obtenu après l'application de ces règles avec des erreurs interprétables et compréhensibles :

Listing 1.2 : Liste des erreurs de l'apprenant à partir des opérations de modification du Listing 1.1

```
('MISSING_CALL_INSTRUCTION', 'BAS', 'Module > For > Body: > Call: bas'),
('MISSING_CONST_VALUE', '1', 'Module > For > Body: > Call: bas > Const: 1')
('UNNECESSARY_CALL_STATEMENT', 'BAS', 'Module > Call: bas')
('UNNECESSARY_CONST_VALUE', '2', 'Module > Call: bas > Const: 2')
('CONST_VALUE_MISMATCH', 'Const: 4', 'Const: 3', 'Module > For > Condition: > Call: range > Const: 4')
```

5.2 Prise en compte de l'apprenant

L'identification des erreurs dans un code, telle que présentée dans la section précédente, donne une information ponctuelle sur une partie de l'activité de l'apprenant, elle n'est cependant pas nécessairement représentative de son profil. Nous intégrons donc, dans la définition de l'état de l'environnement du système de décision, un profil de l'apprenant. L'approche mise en œuvre est la définition d'un *profil des erreurs de l'apprenant* (ses limites sont discutées dans la conclusion). Nous nous inspirons des approches *Buggy Model* [6] en définissant un vecteur qui enregistre les fréquences pondérées des erreurs réalisées par l'apprenant. La pondération est faite selon des considérations temporelles (les erreurs récentes sont plus pondérées que celles des soumissions antérieures). Ce *profil des erreurs de l'apprenant* est mis à jour grâce à la méthode de détection des erreurs présentée section 5.1. Il constitue un composant important de l'état de l'environnement que perçoit l'agent RL et qui lui permet d'apprendre par essais-erreurs (explore-exploit tradeoff) une politique optimale de décision des rétroactions.

6 Définition de l'ensemble des actions du MDP : modélisation et production des rétroactions

Les travaux relatifs aux rétroactions dans l'apprentissage sont nombreux. Concernant plus spécifiquement les rétroactions dans le contexte de l'apprentissage de

[7] urlhttps ://pypi.org/project/asterrdetection/

la programmation nous pouvons citer le travail de Keunig et al. [14], qui propose une revue systématique de la littérature sur la génération automatique de rétroactions pour des exercices de programmation, et l'utilisation des travaux de Narciss [18] par Branthöme dans le cadre de l'environnement Pyrates [4].

Différents travaux ([2], [20]) visent actuellement à la mise en place de bot permettant de générer de manière dynamique des rétroactions à l'aide de LLM. Nous avons fait le choix de ne pas suivre cette voie principalement pour pouvoir contrôler expérimentalement nos choix. En particulier vis à vis du cadre théorique proposé et du modèle de décision que nous souhaitons mettre en place. En effet, la génération dynamique avec LLM a comme conséquence que la décision de la rétroaction est une boîte noire, alors que c'est une de nos questions de recherche. Nous envisageons dans une itération future, quand nous aurons validé expérimentalement cette étape de génération, de produire des rétroactions à partir des LLM tout en utilisant le modèle proposé pour ensuite disposer d'un lot de rétroactions statiques au service de la décision (voir Fig. 7) permettant le passage à l'échelle de notre approche. Dans notre système les rétroactions sont donc prédéfinies et annotées selon différentes caractéristiques, y compris dans l'objectif de pouvoir contrôler l'effet des différentes caractéristiques.

6.1 Modèle de rétroactions

Notre modèle de rétroaction est présenté Fig. 7. Nous détaillons tout d'abord les différents éléments du modèle et leur raison d'être, puis nous l'illustrons avec quelques exemples de rétroactions.

Nos rétroactions sont en lien avec la tâche à réaliser. Le premier constituant est donc un élément relatif à la tâche, il porte sur une *technique* (i.e. "comment réaliser la tâche", voir Fig. 10) ou sur le *logos* (i.e. "savoir à mobiliser pour réaliser la tâche", voir Fig. 8).

Lorsqu'une rétroaction est proposée à l'apprenant il est confronté à plusieurs tâches : 1) comprendre son contenu ; 2) interpréter la ou les raisons pour lesquelles elle lui est proposée ; 3) l'exploiter pour agir sur son travail.

C'est l'intention d'étayage [22] de ces tâches qui nous amène à augmenter le contenu "technique / logos" des éléments suivants.

Pour aider à comprendre et exploiter le contenu de la rétroaction, nous pouvons :

- l'enrichir avec un exemple qui illustre le savoir ou une partie de la technique, voir Fig. 9 et Fig. 10
- pointer explicitement une erreur fréquente associée à la technique ou au logos évoqué, voir Fig. 8 et Fig. 9
- instancier le contenu de tout ou partie de la rétroaction à la tâche à résoudre en intégrant à son contenu des éléments de l'énoncé ou des éléments de la réponse de l'apprenant
- ajouter une phrase qui indique le statut d'un contenu et/ou la raison pour laquelle ce contenu est proposé

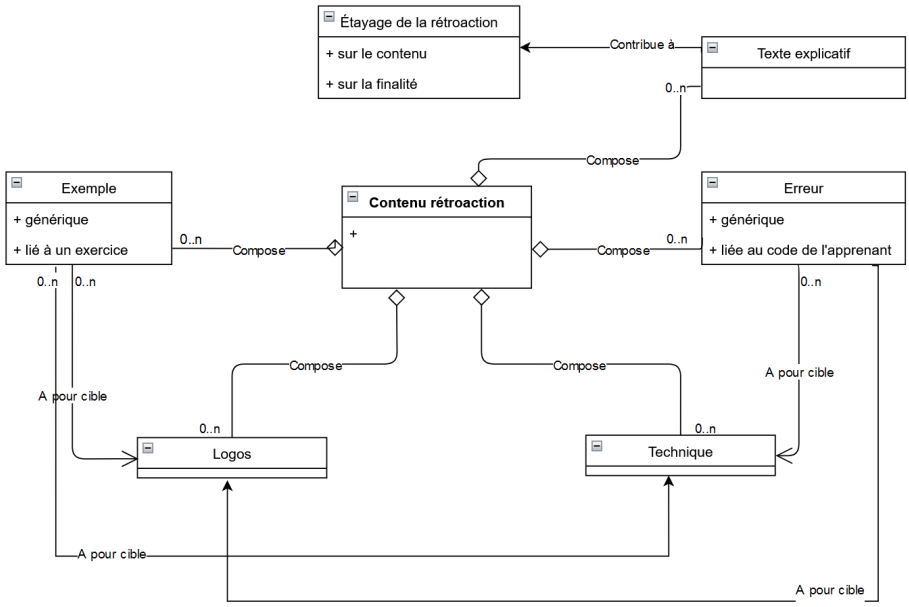


FIGURE 7 : Modélisation des rétroactions dans AlgoPython

Une boucle est une structure qui permet de répéter une ou plusieurs instructions.

Fais attention en particulier au nombre de répétitions.

FIGURE 8 : Rétroaction du type : logos avec l'erreur "For Loop Incorrect number of iteration" pointée

6.2 Production et annotation des rétroactions

Une rétroaction est un assemblage de différents éléments identifiés Fig. 7. Pour produire les rétroactions relatives à la tâche, nous définissons la règle suivante : une rétroaction contient *a minima* un élément de type technique ou un élément de type logos (pour contrôler l'expérimentation et pour évaluer l'effet du contenu sur un apprenant nous faisons le choix, dans un premier temps, de ne pas produire de rétroaction contenant simultanément logos et technique).

Ce contenu peut être complété par les différents éléments présentés dans la section 6.1. Pour une série d'exercices d'AlgoPython nous produisons trois catégories de rétroactions : celles relatives au thème de la série en général ; celles relatives à une erreur spécifique de l'apprenant et celles spécifiques à un type d'énoncé.

Pour qu'une rétroaction puisse être décidée il est nécessaire qu'elle soit décrite. Cette description est basée d'une part sur les éléments décrits dans la Fig. 7 auxquels sont ajoutés les informations suivantes :

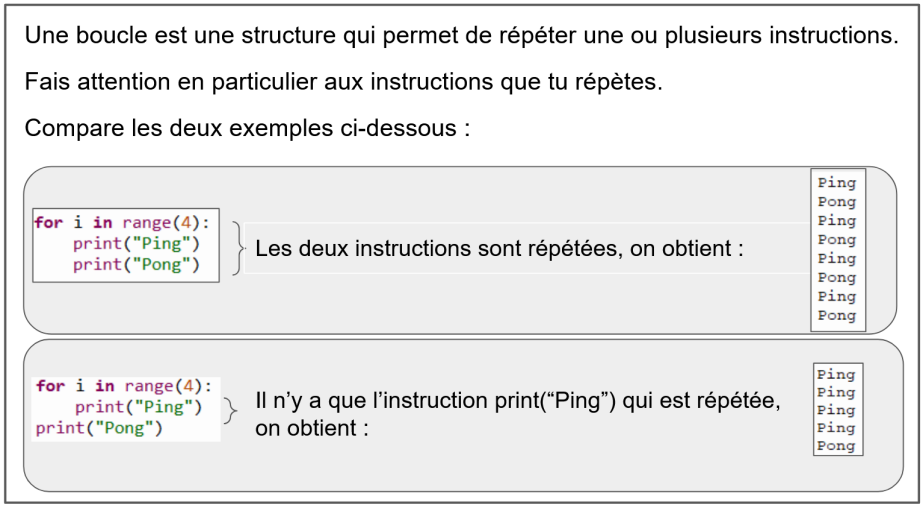


FIGURE 9 : Rétroaction du type : logos avec l'erreur "For loop body mismatch" pointée et un exemple non lié à un exercice spécifique

- Le ou les types de tâches auxquels associer la rétroaction.
- Le cas échéant, le ou les exercices spécifiques auxquels elle est destinée.
- La ou les erreurs (voir section 5.1) pour lesquelles elle fait sens.

7 Preuve de concept sur l'adéquation de l'approche RL pour la décision des rétroactions

Nous avons mené un premier travail exploratoire pour vérifier qu'un framework RL est adéquat pour l'apprentissage d'une politique optimale de décision des rétroactions avant de le déployer dans la plateforme AlgoPython. Pour cela, nous avons lancé des simulations sur des données synthétiques. Dans le cadre de ces simulations, nous avons fait des choix pour la génération des données simulées :

- l'état de l'environnement est restreint à des informations sur la tâche en cours, aux erreurs de la production en cours et à l'état de l'apprenant,
- l'état d'un apprenant est simplifié et est défini uniquement par son profil des erreurs (généré de manière aléatoire),
- les différentes erreurs d'un profil des erreurs sont indépendantes les unes des autres, c'est à dire, l'apparition d'une erreur n'influence pas l'apparition d'une autre erreur,
- on considère 5 types d'apprenants différents, un apprenant appartient à un type si la majorité des fréquences de ses erreurs appartient à un intervalle prédéfini.

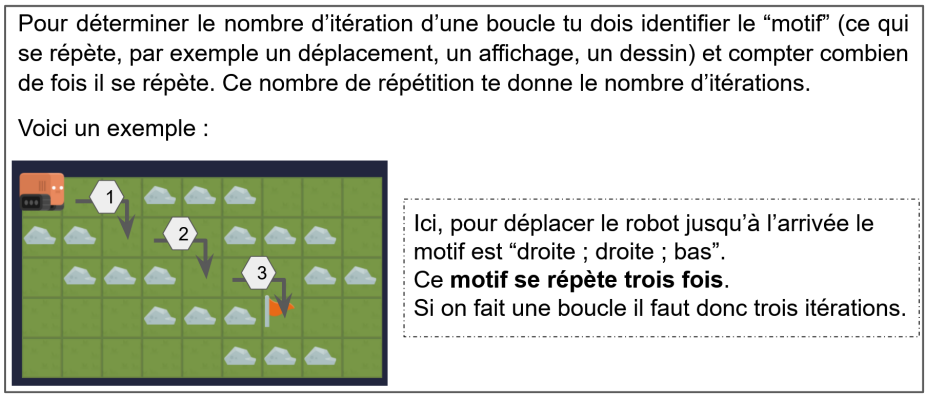


FIGURE 10 : Rétroaction du type : technique et exemple lié à un exercice spécifique

En plus de ces choix de simulation, nous générons les données selon la politique de décision (*Behavior policy*) suivante : on associe à chaque type d'apprenants simulé une rétroaction adaptée (voir section 6). Cette dernière, lorsqu'elle est décidée pour l'apprenant simulé à l'état S_t lui permet de corriger son erreur à l'état S_{t+1}, c'est à dire que la fréquence de cette erreur dans le profil des erreurs de l'apprenant diminue.

Bien entendu, ces choix de génération de données simulées ne correspondent aucunement à la réalité de données réelles issues d'expérimentations écologiques avec de vrais apprenants. Elles permettent néanmoins de vérifier l'efficacité des approches RL à apprendre une politique de décision cohérente avec les données simulées.

La question de recherche à laquelle nous voulons répondre à travers ces simulations, qui rejoint la QR2 (Voir section 2), est : **Est-ce qu'un modèle RL peut apprendre à décider la rétroaction la plus adaptée pour chaque type d'apprenants simulé ?**

Types d'apprenant simulés Nous avons défini 5 types d'apprenants bien distincts. Un apprenant appartient à un type i si la plupart des fréquences de ses erreurs appartiennent à l'intervalle $[1-0.2*(i), 1-0.2*(i-1)]$ sachant que la fréquence d'une erreur est une valeur dans $[0, 1]$. Concrètement, les apprenants du type 1 sont les apprenants qui font le plus d'erreurs et ceux du type 5 sont ceux qui font le moins d'erreurs.

Types de rétroactions Étant dans une cadre simulé, nous avons simplifié le problème d'adaptation d'une rétroaction à un apprenant comme suit : on associe à chaque type d'apprenants une et une seule rétroaction adaptée. Ainsi, toutes les autres rétroactions ne sont pas adaptées.

Fonction action-valeur ou récompenses cumulatives La fonction de récompense R a été définie comme suit :

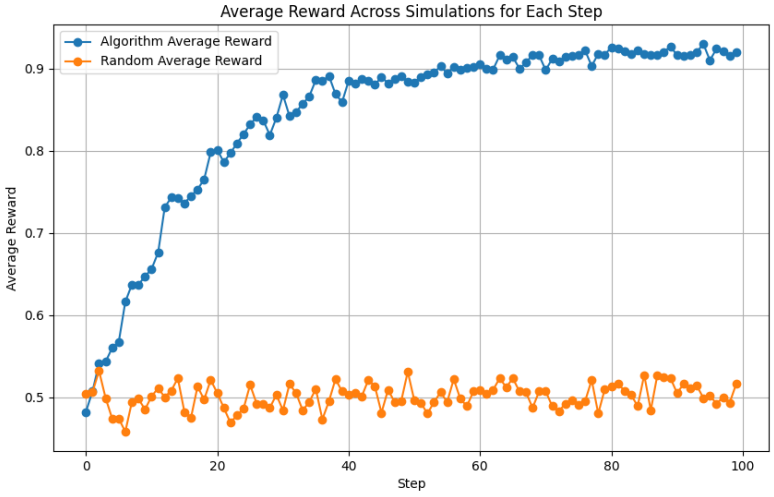
$$R(s_t, a_t) = \begin{cases} 1 & \text{Si la rétroaction } a_t \text{ est celle adaptée au type de l'apprenant dont le profil des erreurs est inclus dans } s_t \\ 0 & \text{Sinon} \end{cases}$$

De manière itérative, l'agent RL calcule les récompenses cumulatives reçues pour chaque décision de rétroaction a à un état s comme suit (t étant le nombre d'étapes d'apprentissage) :

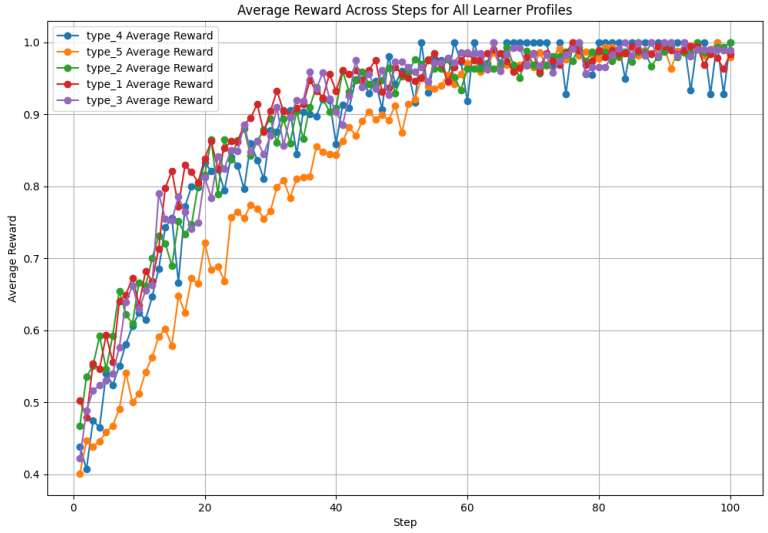
$$\begin{aligned} Q_{t+1}(s,a) &= \frac{1}{t}\left(r_1(s,a) + r_2(s,a) + \cdots + r_t(s,a)\right) \\ &= \frac{1}{t}\left((t-1)Q_t(s,a) + r_t(s,a)\right) \\ &= \frac{1}{t}\left(tQ_t(s,a) + r_t(s,a) - Q_t(s,a)\right) \\ &= Q_t(s,a) + \frac{1}{t}\left(r_t(s,a) - Q_t(s,a)\right) \end{aligned}$$

C'est la mise à jour itérative de cette fonction de valeur Q qui permet à l'agent RL d'apprendre à décider la bonne rétroaction à chaque état de l'environnement, c'est à dire, la rétroaction qui maximise les récompenses reçues. Dans cette simulation, une récompense est reçue lorsque la rétroaction permet de corriger une ou plusieurs erreurs de l'apprenant simulé.

Modèles RL utilisés Pour répondre à notre (QR2.1), nous avons testé deux modèles RL : Linear Contextual Multi-Armed Bandits avec Thompson Sampling ; Neural Contextual Multi-Armed Bandits avec Thompson Sampling. Nous présentons ici les résultats du modèle *Linear Contextual Multi-Armed Bandits avec Thompson Sampling* mais les résultats obtenus avec l'autre modèle testé sont similaires. Pour entraîner ce modèle sur nos données simulées, nous avons lancé 1000 simulations. Dans chaque simulation, nous avons généré 100 états de manière aléatoire en veillant à ce que chaque type d'apprenant soit représenté de manière équitable. Puis, nous avons calculé la moyenne des récompenses sur l'ensemble des simulations. Fig. 11(a) (respectivement la Fig. 11(b)) montre l'évolution de la moyenne des récompenses pour l'ensemble des types d'apprenant (respectivement par type d'apprenants) au fur et à mesure de l'apprentissage. On constate que la moyennes des récompenses augmente et se stabilise rapidement autour de 10 pas d'apprentissage. Pour mieux apprécier ce résultat, nous avons représenté par la courbe orange la moyenne des récompenses obtenues avec une politique de décision aléatoire sur les 100 pas de temps (Fig. 11(a)). On peut constater que la politique RL performe beaucoup mieux que la politique aléatoire et que la politique apprise (target policy) et la politique de génération des données (behavior policy) sont très proches. Ainsi, les résultats de ces premières simulations montrent qu'il est possible d'apprendre à un modèle RL la



(a) Moyenne des récompenses



(b) Moyenne des récompenses par type d'apprenants

FIGURE 11 : Linear Contextual Multi-Armed Bandits avec Thompson Sampling

politique de décision qui a été utilisée pour la génération des données. Ces résultats renforcent notre hypothèse sur l'adéquation des modèles RL à la décision des rétroactions adaptées.

8 Conclusion et perspectives

Dans cet article nous avons présenté les choix et premiers éléments mis en place pour permettre de décider des rétroactions épistémiques dans l'environnement AlgoPython. Pour répondre à la première question de recherche (QR1), nous avons proposé un modèle de rétroactions permettant à la fois leur production et leur annotation.

En ce qui concerne la deuxième question de recherche (QR2), nous avons cherché à y répondre en explorant une modélisation de la décision de type RL avec une définition de l'état de l'environnement, notamment basée sur les erreurs de l'apprenant. À cette occasion nous avons aussi produit une méthode d'analyse des codes soumis avec la mise à disposition d'un package Python original [8]. Enfin, nous avons fait une première évaluation exploratoire avec des données simulées afin de vérifier que le modèle converge de façon satisfaisante.

Ces travaux ont été réalisés en exploitant le référentiel de types de tâches de programmation proposé dans [13] qui a fournit un socle commun pour aborder nos différentes questions.

Nous engageons maintenant une première expérimentation, avec une dizaine d'enseignants, avec un premier déploiement de l'algorithme qui va travailler sur la décision des 75 rétroactions produites pour les séries *instructions*, *boucle For* et *variables* d'AlgoPython. Cette expérimentation en conditions réelles nous permettra d'obtenir un premier ensemble de données relatives à la version d'AlgoPython augmentée par les rétroactions.

Le travail de modélisation des rétroactions va être enrichi et exploité à l'aide de trois travaux, menés en parallèle, qui débutent :

- faire évaluer par des enseignants et des élèves le contenu des rétroactions produites
- explorer l'exploitation d'IA génératives avec une approche RAG et chain-of-tought [12] pour permettre une production (semi)automatisée de contenus de rétroactions
- améliorer le dispositif d'annotation d'erreur présenté dans la section 4.1

Ils doivent permettre d'améliorer le contenu des rétroactions et la connaissance de l'apprenant permettant à l'algorithme de décision des rétroactions de travailler dans des conditions optimisées du point de vue de l'apprentissage humain.

Au niveau du modèle décisionnel, la suite du travail porte d'une part sur la fonction de récompense (section 4) et d'autre part sur la modélisation de l'apprenant (section 5.2) pour mieux informer l'agent RL. En effet, actuellement

[8] https ://pypi.org/project/asterrdetection/

nous modélisons uniquement le profil des erreurs des apprenants qui n'informe pas sur l'évolution de leur niveau de maîtrise des connaissances. Nous souhaitons traiter cette limite en explorant des modèles de traçage des connaissances tel que le modèle BKT (Bayesian Knowledge Tracing) [31].

Références

1. Aleven, V., McLaughlin, E.A., Glenn, R.A., Koedinger, K.R. : Instruction based on adaptive learning technologies. Handbook of research on learning and instruction **2**, 522–560 (2016)
2. Balse, R., Valaboju, B., Singhal, S., Warriem, J.M., Prasad, P. : Investigating the potential of gpt-3 in providing feedback for programming assessments. In : Proceedings of the 2023 Conference on Innovation and Technology in Computer Science Education V. 1. p. 292–298. ITiCSE 2023, Association for Computing Machinery, New York, NY, USA (2023). https://doi.org/10.1145/3587102.3588852, https ://doi.org/10.1145/3587102.3588852
3. Bassen, J., Balaji, B., Schaarschmidt, M., Thille, C., Painter, J., Zimmaro, D., Games, A., Fast, E., Mitchell, J.C. : Reinforcement learning for the adaptive scheduling of educational activities. In : Proceedings of the 2020 CHI conference on human factors in computing systems. pp. 1–12 (2020)
4. Branthome, M. : Apprentissage de la programmation informatique : analyses et ressources pour accompagner la transition collège-lycée. PhD Thesis, Université de Bretagne occidentale-Brest (2023), https ://theses.hal.science/tel-04397329/
5. Brooks, C., Carroll, A., Gillies, R.M., Hattie, J. : A matrix of feedback for learning. Australian Journal of Teacher Education (Online) **44**(4), 14–32 (2019)
6. Brown, J.S., Burton, R.R. : Diagnostic models for procedural bugs in basic mathematical skills. Cognitive science **2**(2), 155–192 (1978)
7. Chevallard, Y. : Concepts fondamentaux de la didactique : perspectives apportées par une approche anthropologique. Recherche en didactique des mathématiques **12**(1), 73–112 (1992)
8. Corbett, A.T., Anderson, J.R. : Knowledge tracing : Modeling the acquisition of procedural knowledge. User modeling and user-adapted interaction **4**, 253–278 (1994)
9. Cunningham, P., Cord, M., Delany, S.J. : Supervised learning. In : Machine learning techniques for multimedia : case studies on organization and retrieval, pp. 21–49. Springer (2008)
10. Ghahramani, Z. : Unsupervised learning. In : Summer school on machine learning, pp. 72–112. Springer (2003)
11. Hattie, J., Timperley, H. : The power of feedback. Review of educational research **77**(1), 81–112 (2007), publisher : Sage Publications Sage CA : Thousand Oaks, CA
12. Jacobs, S., Jaschke, S. : Leveraging Lecture Content for Improved Feedback : Explorations with GPT-4 and Retrieval Augmented Generation. In : 2024 36th International Conference on Software Engineering Education and Training (CSEE&T). pp. 1–5. Germany (2024). https://doi.org/10.1109/CSEET62301.2024.10663001, http ://arxiv.org/abs/2405.06681, arXiv :2405.06681 [cs]

13. Jolivet, S., Dechaux, E., Gobard, A.C., Wang, P. : Construction et exploitation d'un référentiel de types de tâches d'apprentissage de la programmation. In : Iksaal, S., Pélissier, C., Gilliot, J.M., Marzin-Janvier (eds.) Actes de la onzième Conférence sur les Environnements Informatiques pour l'Apprentissage Humain (EIAH). pp. 154–165. Brest (2023), https ://eiah2023.sciencesconf.org/data/pages/actesEIAH2023.pdf
14. Keuning, H., Jeuring, J., Heeren, B. : A Systematic Literature Review of Automated Feedback Generation for Programming Exercises. ACM Transactions on Computing Education **19**(1), 1–43 (Jan 2019). https://doi.org/10.1145/3231711, https ://dl.acm.org/doi/10.1145/3231711
15. Van der Kleij, F.M., Feskens, R.C.W., Eggen, T.J.H.M. : Effects of Feedback in a Computer-Based Learning Environment on Students' Learning Outcomes : A Meta-Analysis. Review of Educational Research **85**(4), 475–511 (Dec 2015). https://doi.org/10.3102/0034654314564881, http ://journals.sagepub.com/doi/10.3102/0034654314564881
16. Lopez, M.I., Luna, J.M., Romero, C., Ventura, S. : Classification via clustering for predicting final marks based on student participation in forums. International Educational Data Mining Society (2012)
17. Millán, E., Loboda, T., Pérez-De-La-Cruz, J.L. : Bayesian networks for student model engineering. Computers & Education **55**(4), 1663–1683 (2010)
18. Narciss, S. : Designing and Evaluating Tutoring Feedback Strategies for digital learning environments on the basis of the Interactive Tutoring Feedback Model. Digital Education Review **23** (2013)
19. Nogry, S., Jean-Daubias, S., Ollagnier-Beldame, M. : Évaluation des EIAH : une nécessaire diversité des méthodes. In : Acte du colloque "Technologies de l'Information et de la Connaissance dans l'Enseignement Supérieur et l'Industrie". pp. 265–271. Compiègne, France (2004), https ://edutice.archives-ouvertes.fr/edutice-00000729/document
20. Pankiewicz, M., Baker, R. : Large language models (gpt) for automating feedback on programming assignments. In : 31st International Conference on Computers in Education (12 2023)
21. Qian, Y., Lehman, J. : Students' Misconceptions and Other Difficulties in Introductory Programming : A Literature Review. ACM Transactions on Computing Education **18**(1), 1–24 (Dec 2017). https://doi.org/10.1145/3077618, https ://dl.acm.org/doi/10.1145/3077618
22. Quintana, C., Reiser, B.J., Davis, E.A., Krajcik, J., Fretz, E., Duncan, R.G., Kyza, E., Edelson, D., Soloway, E. : A Scaffolding Design Framework for Software to Support Science Inquiry. Journal of the Learning Sciences **13**(3), 337–386 (Jul 2004). https://doi.org/10.1207/s15327809jls1303_4, http ://www.tandfonline.com/doi/abs/10.1207/s15327809jls1303_4
23. Shute, V.J. : Focus on Formative Feedback. Review of Educational Research **78**(1), 153–189 (Mar 2008). https://doi.org/10.3102/0034654307313795
24. Singh, R., Gulwani, S., Solar-Lezama, A. : Automated feedback generation for introductory programming assignments. In : Proceedings of the 34th ACM SIGPLAN Conference on Programming Language Design and Implementation. pp. 15–26. ACM, Seattle Washington USA (Jun 2013). https://doi.org/10.1145/2491956.2462195
25. Small, M., Lin, A. : Instructional feedback in mathematics. In : Lipnevich, A., Smith, J. (eds.) The Cambridge handbook of instructional feedback, pp. 169–190. Cambridge University Press (2018)

26. Sutton, R.S., Barto, A.G. : Reinforcement learning : An introduction. MIT press (2018)
27. Vandewaetere, M., Desmet, P., Clarebout, G. : The contribution of learner characteristics in the development of computer-based adaptive learning environments. Computers in Human Behavior **27**(1), 118–130 (2011)
28. Vie, J.J., Popineau, F., Bruillard, É., Bourda, Y. : A review of recent advances in adaptive assessment. Learning analytics : Fundaments, applications, and trends : A view of the current state of the art to enhance e-learning pp. 113–142 (2017)
29. Walkington, C.A. : Using adaptive learning technologies to personalize instruction to student interests : The impact of relevant contexts on performance and learning outcomes. Journal of educational psychology **105**(4), 932 (2013)
30. Wisniewski, B., Zierer, K., Hattie, J. : The power of feedback revisited : A meta-analysis of educational feedback research. Frontiers in Psychology **10**, 3087 (2020). https://doi.org/https://doi.org/10.3389/fpsyg.2019.03087, publisher : Frontiers
31. Yudelson, M.V., Koedinger, K.R., Gordon, G.J. : Individualized bayesian knowledge tracing models. In : Artificial Intelligence in Education : 16th International Conference, AIED 2013, Memphis, TN, USA, July 9-13, 2013. Proceedings 16. pp. 171–180. Springer (2013)
32. Zhang, K., Shasha, D. : Simple fast algorithms for the editing distance between trees and related problems. SIAM journal on computing **18**(6), 1245–1262 (1989), publisher : SIAM

Analyse des profils d'utilisation de ChatGPT chez les étudiants de 12 à 29 ans en Fédération Wallonie-Bruxelles

Audrey Kumps[1], Melody Di Pietro[2], Sabrin Housni[1], Bruno De Lièvre[1]

[1] Université de Mons
audrey.kumps, sabrin.housni, bruno.delievre{@umons.ac.be}

[2] Université Libre de Bruxelles
melody.di.petro@ulb.be

Résumé. L'étude examine les usages de ChatGPT chez les étudiants âgés de 12 à 29 ans en Fédération Wallonie-Bruxelles, en identifiant quatre profils d'utilisateurs : chercheurs d'informations, créateurs de contenu, utilisateurs limités et utilisateurs hybrides. Ces profils illustrent une diversité d'usages allant de la recherche d'informations à la production de contenus, tout en révélant des disparités dans les perceptions de l'utilité et de l'utilisabilité de l'outil. Parmi les répondants, 70,95 % ont déjà utilisé ChatGPT, tandis que 29,05 % n'y ont jamais eu recours. Les résultats soulignent l'importance de former les étudiants à une utilisation critique et éthique de l'IAG, notamment en ce qui concerne la validation des informations générées et l'optimisation des prompts. Ils mettent également en évidence la nécessité d'un accompagnement différencié. Enfin, des pistes éducatives spécifiques sont proposées, adaptées aux différents profils, afin de maximiser les bénéfices de l'outil tout en réduisant les inégalités d'usage.

Mots-clés : Intelligence artificielle générative, ChatGPT, profils d'utilisateurs, vulnérabilité numérique, fracture numérique

Abstract. The study explores the use of ChatGPT among students aged 12 to 29 in the Wallonia-Brussels Federation, identifying four user profiles: information seekers, content creators, occasional users, and versatile users. These profiles reflect a range of uses, from information retrieval to content production, while also revealing disparities in perceptions of the tool's utility and usability. Among respondents, 70.95% have used ChatGPT, whereas 29.05% have never engaged with it. The findings emphasize the importance of training students in critical and ethical use of generative AI, particularly in validating generated information and optimizing prompts. They also highlight the need for differentiated support. Finally, specific educational strategies are proposed, tailored to the different profiles to maximize the tool's benefits while reducing inequalities in usage.

Keywords: Generative articial Intelligence, ChatGPT, User profiles, Digital vulnerability, Digital Divide

1 Introduction

L'année 2023 a marqué un tournant dans l'appropriation des intelligences artificielles génératives (IAG), notamment avec l'essor de ChatGPT, modèle de langage développé par OpenAI. Ce progrès technologique a rapidement trouvé un écho dans les systèmes éducatifs à travers le monde, touchant différents niveaux de l'enseignement (Madden et al., 2024). Toutefois, l'intégration des intelligences artificielles génératives (IAG) dans l'éducation soulève des enjeux majeurs d'ordre éthique, réglementaire et pédagogique (Lobet et al., 2023). Ces technologies, comme ChatGPT, stimulent des débats autour de problématiques telles que la désinformation, l'intégrité académique, la pensée critique, les enjeux environnementaux et la sécurité des données, mettant en lumière la nécessité de définir des usages responsables (An Ngo, 2023 ; Rudolph et al., 2023 ; Fengchun & Wayne, 2024). En parallèle, leur potentiel pédagogique ne peut être ignoré : elles offrent des opportunités d'innovation et d'amélioration des apprentissages, notamment en répondant aux besoins d'efficacité et d'autonomie des étudiants (Fengchun & Wayne, 2024). Par exemple, ChatGPT permet aux apprenants d'explorer des sujets en profondeur grâce à des réponses instantanées et adaptées (Cotton et al., 2023 ; Lelong, 2023).

Dans ce contexte, les recherches montrent une adoption croissante de ChatGPT, tant par les étudiants du secondaire que ceux de l'enseignement supérieur, qui l'utilisent pour diversifier et approfondir leurs apprentissages (Madden et al., 2024 ; Goudey et al., 2024 ; Pôle Léonard de Vinci, 2024). Toutefois, cette popularité soulève des enjeux, notamment la nécessité d'accompagner enseignants et apprenants dans une utilisation éclairée et responsable, tant sur le plan technologique qu'éthique, afin de maximiser les bénéfices tout en prévenant les abus (Jishnu et al., 2023 ; Abbasi et al., 2024). Malgré cet engouement, les pratiques spécifiques d'utilisation de ChatGPT par les étudiants et leurs effets réels sur l'apprentissage restent encore peu étudiés (Sullivan et al., 2023), particulièrement dans le contexte de la Fédération Wallonie-Bruxelles.

C'est dans ce cadre que s'inscrit la présente recherche, qui propose une contribution en établissant une typologie empirique des profils d'utilisation de ChatGPT, fondée sur les dimensions fonctionnelles d'usage (informationnelle et productive), combinée à l'analyse des perceptions d'utilité et d'utilisabilité de l'outil. En explorant à la fois le secondaire et le supérieur, cette étude apporte un éclairage sur l'hétérogénéité des pratiques et sur les risques d'inégalités d'appropriation de l'IAG en milieu éducatif francophone. Elle vise à (1) décrire les usages actuels de ChatGPT par les étudiants, et (2) identifier des tendances en matière de besoins de compétences liées aux IAG, afin de proposer des pistes exploratoires pour leur intégration pédagogique. L'approche est principalement quantitative et exploratoire, visant à ouvrir des perspectives d'actions éducatives, sans prétendre fournir une compréhension exhaustive ou qualitative des besoins individuels. La question centrale de cette recherche est la suivante : quels sont les profils d'utilisation de ChatGPT par les étudiants du secondaire et du supérieur en Fédération Wallonie-Bruxelles ?

2 Cadre théorique

2.1 L'IAG pour l'éducation

L'IAG, et notamment ChatGPT (pour "Chat Generative Pre-trained Transformer"), peut représenter une opportunité pour l'éducation grâce à ses capacités à générer du texte en langage naturel. Cet outil repose sur des modèles statistiques et une vaste base de données d'entraînement, lui permettant de produire des textes cohérents, de répondre à des questions et de fournir des explications dans divers domaines (Floridi & Chiriatti, 2020). Ces fonctionnalités en font un atout pour soutenir l'apprentissage, qu'il s'agisse de la recherche d'informations, de la compréhension de concepts complexes ou de l'assistance à la production écrite (*ibid.*). An Ngo (2023) va plus loin en qualifiant ChatGPT de véritable assistant virtuel, capable de proposer un service de tutorat personnalisé grâce à des réponses adaptées aux besoins spécifiques des apprenants, renforçant ainsi son efficacité pédagogique.

Cependant, malgré ses performances, ChatGPT présente des limites importantes liées à son fonctionnement. Comme le rappellent Lobet et al. (2024), celle-ci repose sur des prédictions statistiques, générant les séquences de mots les plus probables dans un contexte donné, mais sans véritable sens de la véracité ni de compréhension du contenu produit. Cette caractéristique soulève des défis pour un usage éducatif responsable, notamment en termes de validation des réponses fournies, de développement de l'esprit critique chez les apprenants et de gestion des biais inhérents aux données d'entraînement. Ces questions, cruciales pour l'intégration de l'IAG dans les pratiques pédagogiques, exigent une réflexion approfondie pour maximiser les bénéfices de ces outils tout en en maîtrisant les risques.

2.2 Utilisation de l'IAG par les apprenants

L'IAG s'intègre de plus en plus dans les pratiques éducatives. Son adoption par les apprenants peut être influencée par des motivations variées et des perceptions qui diffèrent selon l'âge, le contexte académique et les besoins spécifiques (Jishnu et al., 2023 ; Hurd & Orban de Xivry, 2024). Les apprenants recourent à des outils comme ChatGPT pour répondre à des besoins diversifiés, notamment la définition de concepts, la compréhension des cours, la résolution d'exercices et l'exemplification d'idées (Lobet et al., 2023). Jishnu et al. (2023) identifient quatre facteurs motivant cette utilisation : (1) la création de contenu académique, comme les devoirs et travaux ; (2) la recherche d'informations, ChatGPT étant perçu comme une alternative rapide et accessible aux moteurs de recherche traditionnels ; (3) l'attraction pour la nouveauté ; (4) la facilité d'utilisation.

L'efficacité est également un facteur clé expliquant l'adoption de l'IAG par les étudiants. Une enquête du Pôle Léonard de Vinci (2024) révèle que 83 % des répondants estiment que ChatGPT leur permet de gagner du temps, et 65 % considèrent qu'il améliore leur productivité. Madden et al. (2024) confirment cet aspect utilitaire en observant que 53 % des adolescents américains utilisent l'IAG principalement pour l'aide aux devoirs, un usage qui illustre son rôle dans la simplification et l'accélération

des tâches scolaires. En Belgique, l'enquête #Génération 2024 (Hurd & Orban de Xivry, 2024) montre que 36 % des étudiants utilisent l'IAG pour des recherches informationnelles et 28 % pour produire du contenu scolaire. Ces résultats soulignent que l'IAG est perçue comme une solution rapide et efficace pour répondre à des besoins éducatifs variés, optimisant ainsi le temps et les efforts investis.

L'âge joue également un rôle dans les usages de l'IAG. Les élèves du secondaire supérieur[1] sont davantage susceptibles d'utiliser ChatGPT pour leurs devoirs que leurs homologues plus jeunes du degré inférieur[2] (Hurd & Orban de Xivry, 2024). Les étudiants universitaires, quant à eux, utilisent l'IAG pour des tâches plus complexes, telles que le développement d'idées ou l'analyse de données (Hurd & Orban de Xivry, 2024 ; Goudey et al., 2024). Cette évolution dans l'usage, liée à l'âge et au niveau d'enseignement, souligne une modification des attentes et des capacités des apprenants à exploiter l'IAG de manière plus avancée.

Cependant, bien que l'IAG soit globalement perçue de manière positive par les apprenants (Bitzenbauer, 2023), elle suscite également des préoccupations concernant la fiabilité et les biais des données générées. Lozano et Blanco Fontao (2023) insistent sur la nécessité de vérifier les informations fournies par ces outils, un point confirmé par Lobet et al. (2023), qui rapportent que 55 % des étudiants ont relevé des erreurs dans les réponses produites par ChatGPT. Cette situation met en évidence l'importance d'un esprit critique face à l'IAG. Dai et al. (2023) différencient deux formes d'apprentissage liées à l'IAG : apprendre à utiliser l'outil, une compétence technique, et apprendre à apprendre avec, ce qui nécessite le développement de stratégies d'autorégulation et d'un usage éthique.

Face à ces enjeux, plusieurs études soulignent l'importance de former les apprenants à une utilisation critique et éthique des IAG. Madden et al. (2024) montrent que les élèves ayant discuté de l'IAG en classe se sentent mieux préparés à en comprendre les avantages et les limites. De même, les enquêtes du Pôle Léonard de Vinci (2024) et de Goudey et al. (2024) révèlent que les étudiants souhaitent apprendre à mieux maîtriser les prompts et comprendre de manière approfondie les implications de l'IAG dans leur futur professionnel.

2.3 Enseigner l'utilisation de l'IAG

L'intégration de l'IAG dans le processus d'apprentissage nécessite que les apprenants développent un éventail de compétences fondamentales, leur permettant de s'engager de manière efficace tout en adoptant une approche éthique et critique. En premier lieu, il est essentiel que les apprenants acquièrent une littératie en IA (Long & Magerko, 2020), comprenant à la fois la compréhension des principes sous-jacents à son fonctionnement et la capacité à évaluer de manière critique les réponses générées (Preiksaitis & Rose, 2023). Cela implique de comprendre les concepts fondamentaux tels que les algorithmes, le machine learning ou encore les modèles de prédiction. Ces

[1] Correspondant au lycée français
[2] Correspondant au collège français

notions peuvent être introduites à travers des activités interactives ou des comparaisons d'outils numériques. Par exemple, les étudiants peuvent être invités à explorer les différences entre ChatGPT et d'autres outils comme Google (pour la recherche d'informations) ou DeepL (pour la traduction), afin d'identifier leurs spécificités et leurs usages respectifs (CLEMI, 2024). Cette initiation leur permet de mieux définir l'IAG, de se familiariser avec ses principes fondamentaux, et de mieux comprendre ses fonctions ainsi que ses limites dans divers contextes d'utilisation (STEMENTIEL, 2024).

Une fois les bases établies, il faut pouvoir interagir efficacement avec ces outils. La formulation de prompts clairs et adaptés constitue une compétence essentielle pour optimiser les réponses de l'IAG. La qualité des résultats générés dépend largement de la précision et de la clarté des instructions fournies (Ahajjam & Bouabbouz, 2023). Les enseignants peuvent organiser des ateliers pratiques où les apprenants expérimentent la création de prompts, testent diverses formulations et affinent leurs requêtes pour obtenir des résultats optimaux (Lelong, 2023). Une telle démarche leur permet de mieux comprendre les mécanismes de fonctionnement de l'IAG et d'en maximiser l'efficacité (Cabana & Lez, 2023).

En parallèle, l'évaluation critique des informations générées par l'IAG est une compétence incontournable (Preiksaitis & Rose, 2023). Les réponses fournies par des outils comme ChatGPT peuvent comporter des « hallucinations », c'est-à-dire des réponses qui semblent plausibles mais qui sont en réalité incorrectes ou inventées (Li et al. 2023). Il est donc nécessaire de former les apprenants à vérifier la fiabilité et la pertinence des informations en les confrontant à des sources externes fiables (Dai et al., 2023). Une activité pratique efficace pourrait consister à concevoir, avec l'IAG, une grille d'évaluation permettant d'analyser la cohérence des données, la validité des références citées ou encore la pertinence des contenus fournis (Degeer & Kumps, 2022). De plus, l'approche de la lecture latérale (Pavlounis et al., 2021), qui invite les apprenants à comparer les réponses générées avec des articles scientifiques, des sites institutionnels ou à rechercher les sources originelles, peut être intégrée dans les pratiques pédagogiques. Cette méthode encourage une posture réflexive face aux informations et s'avère essentielle dans un monde saturé de données.

Les outils d'IAG offrent un soutien précieux pour des tâches telles que la rédaction, la génération de résumés ou encore le brainstorming d'idées créatives. Cependant, ces outils doivent être utilisés comme des points de départ plutôt que comme des solutions fiables afin de préserver la réflexion personnelle et de maintenir l'intégrité académique dans leurs travaux (Diallo, 2023 ; Rudolph et al., 2023). Il est essentiel que les enseignants forment les apprenants aux limites et aux possibilités des IAG en valorisant la production écrite par les apprenants eux-mêmes tout en intégrant une sensibilisation aux enjeux éthiques, notamment le plagiat, la propriété intellectuelle et les biais algorithmiques. Ces discussions permettent aux apprenants de comprendre que, potentiellement efficaces, ces outils ne remplacent pas la réflexion propre aux individus (Cotton et al., 2023).

3 Méthodologie

Cette recherche se focalise spécifiquement sur ChatGPT, en tant que système d'IAG largement utilisé (Pôle Léonard de Vinci, 2024 ; Fletcher & Nielsen, 2024), justifiant ce choix.

3.1 Collecte de données

Nous avons adopté une approche quantitative via un questionnaire en ligne, conçu avec Microsoft Forms et diffusé par les adresses officielles des établissements répertoriés sur le site de la Fédération Wallonie-Bruxelles. L'enquête, strictement anonyme et conforme aux principes du Règlement Général sur la Protection des Données (RGPD) de l'Union européenne (Demotes-Mainard et al., 2018 ; Plessz, 2020), garantissait la confidentialité des répondants sans collecte de données identifiantes.

3.2 Outil de mesure

Le questionnaire est structuré en quatre sections afin de mettre en avant différents aspects de l'utilisation de ChatGPT :

- Caractérisation de l'échantillon

Cette section recueille des informations démographiques (âge, sexe) et contextuelles (niveau d'enseignement, fréquence d'utilisation des outils numériques en minutes par semaine). Ces données permettent de décrire l'échantillon et d'identifier des tendances selon les caractéristiques des répondants.

- Familiarité avec ChatGPT

Cette section évalue l'expérience préalable des participants avec ChatGPT. Ceux qui ne l'ont jamais utilisé voient leur questionnaire se terminer à ce stade. Pour les utilisateurs de ChatGPT, des questions supplémentaires explorent leur utilisation personnelle et académique, ainsi que leurs perceptions quant à son utilité et son utilisabilité.

- Utilisation effective de ChatGPT

Les participants qui ont déjà utilisé ChatGPT répondent à des items inspirés des travaux de Jishnu et al. (2023), Goudey et al. (2024), ainsi que de Madden et al. (2024). Ces items couvrent des usages variés tels que la rédaction de travaux, la préparation de présentations, ou encore la recherche d'informations (Tableau 1). Les réponses sont évaluées à l'aide d'une échelle de Likert à cinq points, allant de «complètement en désaccord» à «complètement d'accord».

Dans cette section, les participants indiquent également la fréquence de leur utilisation de ChatGPT dans un cadre académique et personnel, sur une échelle de 0 (jamais) à 5 (plusieurs fois par jour).

- Perceptions de l'utilité et de l'utilisabilité de ChatGPT

Cette section explore les représentations des participants concernant ChatGPT, en adaptant l'échelle de perception technologique de Davis (1989) aux outils d'IAG (Brangier & Hammes, 2007). Les dimensions mesurées incluent l'utilité (évaluant l'effet de l'outil sur les performances académiques et professionnelles) et l'utilisabilité (mesurant des aspects tels que la facilité d'utilisation, la clarté des interactions et la

flexibilité). Chaque item est évalué sur une échelle de Likert à cinq points, de « pas du tout d'accord » à « tout à fait d'accord » (Tableau 2).

Tableau 1. Items relatifs aux usages de ChatGPT

Items J'utilise ChatGPT pour…	**Sources**
rédiger mes devoirs	Jishnu et al. (2023), Goudey et al. (2024), Madden et al. (2024)
préparer mes présentations orales	Jishnu et al. (2023), Goudey et al. (2024), Madden et al. (2024)
écrire mes travaux de recherche scientifique	Jishnu et al. (2023), Goudey et al. (2024)
écrire une thèse/une dissertation/un travail	Jishnu et al. (2023), Goudey et al. (2024), Madden et al. (2024)
rédiger mes notes de cours	Jishnu et al. (2023), Goudey et al. (2024)
écrire un résumé	Jishnu et al. (2023), Goudey et al. (2024)
obtenir de l'information sur un sujet pour le définir	Jishnu et al. (2023), Lobet et al. (2023), Goudey et al. (2024)
recueillir des informations pour mon développement personnel et académique	Jishnu et al. (2023)
obtenir des informations pour un cours (informations supplémentaires ou meilleure compréhension)	Jishnu et al. (2023), Lobet et al. (2023), Goudey et al. (2024)
valider ou vérifier une information	Jishnu et al. (2023)
obtenir des informations sur la planification et l'organisation	Jishnu et al. (2023), Pole Léonard de Vinci (2024)
obtenir de l'information sur un sujet (pour mieux le comprendre)	Jishnu et al. (2023), Lobet et al. (2023), Goudey et al. (2024)
me préparer aux examens	Jishnu et al. (2023)

Tableau 2. Items relatifs à la perception d'utilité et d'utilisabilité de ChatGPT

Variables	**Catégories**	**ChatGPT est un outil …**
Apprentissage	Utilisabilité	pour lequel il m'a été facile d'apprendre le fonctionnement
Facilité		qui m'a semblé facile à utiliser
Clarté et compréhension		avec lequel l'interaction est claire et compréhensible
Contrôlabilité		qui est facile à utiliser pour concevoir une carte conceptuelle
Habileté		avec lequel il m'a été facile de devenir compétent
Flexibilité		avec lequel je peux interagir en souplesse
Rapidité	Utilité	qui me permet de réaliser mes tâches plus rapidement
Performance		qui améliore la qualité de mon travail
Productivité		qui augmente ma productivité
Facilité		qui facilite l'accomplissement de ma tâche
Efficacité		qui améliore l'efficacité de mon apprentissage
Transfert		qui me sera utile dans mon contexte professionnel

3.3 Échantillon

L'échantillon de l'étude est composé de 864 participants, répartis entre l'enseignement secondaire (41 % ; N=351) et l'enseignement supérieur (59 % ; N=513). L'âge moyen des répondants est de 20,98 ans (σ=6,93), et comprend une majorité de femmes (61 % ; N=529). Les répondants rapportent une utilisation quotidienne des outils numériques, avec une moyenne de 3h50 par jour dans un cadre académique pour les étudiants du supérieur, contre 1h57 pour les élèves du secondaire.

Les participants inscrits dans l'enseignement secondaire, âgés en moyenne de 15,98 ans (σ=2,27), proviennent soit de l'enseignement général (64,95 % ; N=228), technique (20,22 % ; N=71) ou professionnel (14, 81% ; N=52). Les options les plus populaires incluent les sciences appliquées (43 % ; N=152) et les langues (14 % ; N=51).

Les participants inscrits dans l'enseignement supérieur, âgés en moyenne de 24,4 ans (σ=6,97), sont soit inscrits en bachelier (82,06 % ; N=421), en master (17,54 % ; N=90) ou au doctorat (0,39 % ; N=2) et suivent principalement des études en sciences politiques, économiques ou psychologiques. Plus de la moitié (52,63 % ; N=270) sont inscrits en Haute École, 23,59 % (N=121) à l'université, et 23,39 % (N=120) en école de promotion sociale. Deux étudiants sont inscrits en école supérieure des arts.

4 Analyse des résultats

4.1 Utilisation de ChatGPT

Les résultats de l'enquête révèlent qu'au sein des 864 répondants, 29,05 % (n=251) déclarent ne jamais avoir utilisé ChatGPT, tandis que 70,95 % (n=613) affirment l'avoir déjà utilisé.

Caractéristiques des non-utilisateurs

Les non-utilisateurs de ChatGPT (n=251) présentent un profil distinct : ils sont en moyenne plus jeunes (19,99 ans ; σ = 7,90) que les utilisateurs (21,39 ans ; σ = 6,46) (U = 57 660,5 ; p<.001) et sont majoritairement issus de l'enseignement secondaire, comme le révèle une relation significative entre leur niveau d'enseignement et l'utilisation de ChatGPT (X^2 = 31,924 ; p <.001).

En outre, leur utilisation des outils numériques est significativement plus restreinte, tant dans un contexte académique (2h40/semaine contre 3h13 pour les utilisateurs) que personnel (3h35/semaine contre 4h01), avec des différences significatives (U=63300,5 ; p<.001 en contexte éducatif ; U=68353 ; p=.009 en contexte personnel). Ces caractéristiques suggèrent que les non-utilisateurs sont moins exposés ou moins enclins à exploiter les outils numériques, ce qui pourrait refléter des besoins ou des habitudes différents, propres à leur profil.

Caractéristiques et usages des utilisateurs

Outre les caractéristiques mentionnées précédemment, les utilisateurs de ChatGPT (N = 613) se distinguent par une fréquence d'utilisation significativement plus élevée dans un contexte académique (M = 3,95) que dans un cadre personnel (M = 3,01) (W = 13.489 ; p < .001), sur une échelle allant de 0 (jamais) à 5 (plusieurs fois par jour).

Leurs usages se diversifient en fonction des besoins (Fig. 1). Les activités les plus fréquentes incluent la recherche d'informations (« obtenir de l'information sur un

sujet » ; 67,2 % d'accord[3] ; N=412) et la compréhension de sujets complexes (« comprendre certains sujets » ; 64,3 % d'accord ; N=394). Plus de la moitié (53,7 % ; N=329) utilisent également ChatGPT pour rédiger des résumés.

En revanche, certaines utilisations spécifiques restent marginales : seuls 15,7 % (N=96) l'emploient pour rédiger des notes de cours, et 23,3 % (N = 143) l'utilisent pour des tâches liées à l'organisation ou à la prise de décision (« obtenir des informations sur la planification, l'organisation et la prise de décisions »). Ces résultats mettent en lumière des usages variés.

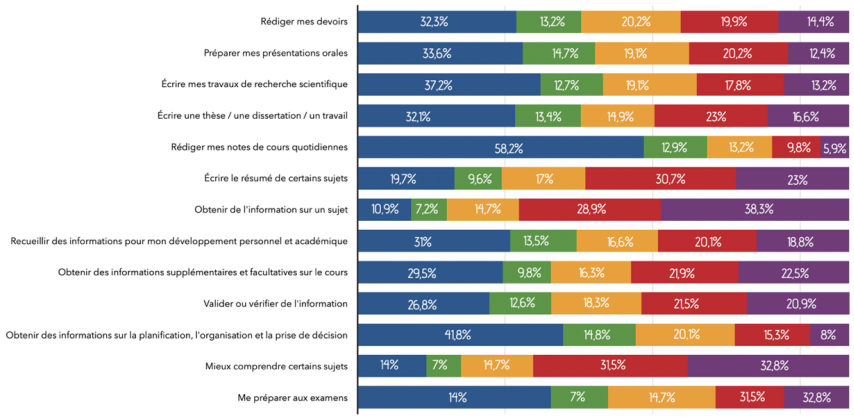


Fig. 1. Utilisation de ChatGPT

Pour approfondir notre analyse, une analyse en composantes principales (ACP), avec rotation promax est effectuée. Une rotation promax est choisie en raison de l'interdépendance théorique des utilisations. Le seuil de saturation significative de chaque facteur est fixé à 0,40. L'ACP a permis de distinguer deux dimensions d'utilisation principales, expliquant 57 % de la variance totale (Tableau 3). La composante 1, que nous nommons « Utilisation informationnelle » (30,2 % de la variance totale) est composée de 7 items. La composante 2, que nous nommons « Utilisation productive » (26,8 % de la variance totale) est composée de 6 items.

[3] L'"accord" désigne les répondants ayant sélectionné les options "d'accord" et "complètement d'accord" sur les échelles de Likert utilisées dans le questionnaire en ligne.

Tableau 3. Analyse en composantes principals

	Composante 1 **Utilisation informationnelle**	Composante 2 **Utilisation productive**
pour obtenir des informations supplémentaires sur le cours.	0.839	
pour mieux comprendre certains sujets.	0.775	
pour recueillir des informations pour mon développement personnel et académique.	0.757	
pour valider ou vérifier de l'information.	0.757	
pour obtenir de l'information sur un sujet.	0.650	
pour obtenir des informations sur la planification, l'organisation et la prise de décision.	0.625	
pour me préparer aux examens.	0.560	
pour rédiger mes notes de cours quotidiennes.		0.445
pour écrire une thèse / une dissertation / un travail.		0.864
pour rédiger mes devoirs.		0.835
pour écrire mes travaux de recherche scientifique		0.813
pour préparer mes présentations orales		0.739
pour écrire le résumé de certains sujets.		0.579

Une analyse par clustering (k-means) basée sur ces composantes a permis d'identifier quatre profils d'utilisateurs parmi les 613 répondants. Les clusters montrent un regroupement cohérent (indice silhouette=0,280) et un bon ajustement global (R^2=0,684).

Les profils identifiés (auxquels nous avons attribué des étiquettes basées sur les dimensions de leur profil) se répartissent comme suit :

1. Les **chercheurs d'informations** (19,58 % ; N=120) : ces utilisateurs exploitent principalement ChatGPT pour la recherche et la validation d'informations, en se concentrant sur son usage informationnel.
2. Les **créateurs de contenu** (22,84 % ; N=140) : ces participants utilisent ChatGPT avant tout pour produire du texte, comme des travaux académiques ou des présentations, tout en recourant peu à ses fonctions de recherche.
3. Les **utilisateurs limités** (42,58 % ; N=261) : représentant le groupe le plus important, ces apprenants interagissent rarement avec l'outil, que ce soit pour s'informer ou générer du contenu, indiquant une utilisation limitée ou sporadique.
4. Les **utilisateurs hybrides** (15,01 % ; N=92) : ce groupe exploite pleinement les capacités de ChatGPT, combinant à la fois la recherche d'informations et la production de contenu, ce qui les distingue par leur usage intensif et varié de l'outil.

La Fig.2, sous forme d'un diagramme en barres, illustre de manière visuelle la relation entre les composantes principales et les clusters.

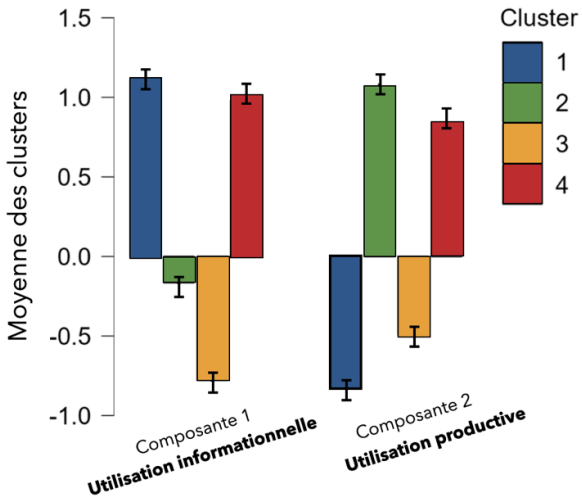


Fig. 2. Clusters d'utilisation de ChatGPT

4.2 Caractéristiques des clusters d'utilisateurs

Variables démographiques

Les analyses des relations entre les clusters d'utilisateurs de ChatGPT et les variables démographiques (âge, niveau d'enseignement et sexe) montrent que ces facteurs ne présentent pas de différences significatives. Plus précisément, l'âge des utilisateurs ne varie pas de manière significative entre les clusters ($W = 0{,}841$; $p = 0{,}840$). De même, aucune relation significative n'a été identifiée entre le niveau d'enseignement (secondaire ou supérieur) et les profils d'utilisation ($x^2 = 2{,}406$; $p = 0{,}492$), ni entre le sexe et les clusters d'utilisateurs ($x^2 = 7{,}922$; $p = 0{,}244$). Ces résultats indiquent que les profils d'utilisateurs ne se distinguent pas au niveau des caractéristiques démographiques.

Perceptions d'utilité et d'utilisabilité de ChatGPT

Les perceptions d'utilisabilité et d'utilité de ChatGPT varient tant au niveau de nos analyses descriptives et qu''inférentielles (Tableau 4) entre les différents clusters identifiés. En particulier, le cluster 3 (**utilisateurs limités**) se distingue par des scores systématiquement inférieurs dans toutes les dimensions, reflétant une perception plus nuancée de l'outil par ces utilisateurs.

Tableau 4. Perception de l'utilité et de l'utilisabilité selon les clusters d'utilisation

Dimension	Cluster 1 Chercheurs d'informations	Cluster 2 Créateurs de contenu	Cluster 3 Utilisateurs occasionnels	Cluster 4 Utilisateurs polyvalents	Statistique	p-valeur
Utilisabilité						
Apprentissage	4,48	4,48	**4,27**	4,62	11,287	0,01
Facilité d'utilisation	4,67	4,56	**4,38**	4,70	17,032	<.001
Clarté et compréhension	4,31	4,25	**4,01**	4,38	14,184	0,003
Contrôlabilité	4,67	4,56	**4,38**	4,70	17,032	<.001
Habileté	3,76	3,87	**3,25**	4,01	41,89	<.001
Flexibilité	4,06	3,96	**3,51**	4,19	39,913	<.001
Utilité						
Rapidité	4,06	4,47	**3,52**	4,5	78,349	<.001
Performance	3,66	3,92	**3,01**	4,18	86,189	<.001
Productivité	3,63	3,85	**2,94**	4,01	72,995	<.001
Facilité	4,00	4,39	**3,56**	4,36	69,982	<.001
Efficacité	3,69	3,46	**2,56**	3,94	111,255	<.001
Transfert	3,97	4,03	**3,12**	4,33	103,286	<.001

Les analyses post-hoc confirment que les « **utilisateurs limités** » présentent des perceptions significativement plus faibles comparées à celles des « **utilisateurs hybrides** » (tous les p sont <0,05). En revanche, les perceptions des « **chercheurs d'informations** », des « **créateurs de contenu** » et des « **utilisateurs hybrides** » sont statistiquement équivalents concernant l'utilité et l'utilisabilité (tous les p sont >0,05).

5 Discussion des résultats

Les résultats obtenus mettent en lumière la diversité des usages de ChatGPT chez les étudiants de la Fédération Wallonie-Bruxelles, tout en identifiant des profils distincts d'utilisateurs. Cette étude confirme les observations préalables de la littérature sur l'intégration croissante des outils d'IAG auprès des apprenants (Jishnu et al., 2023 ; Lobet et al., 2023). L'analyse montre que 70,95 % des répondants ont déjà utilisé ChatGPT, avec des usages variés répartis entre la recherche d'informations, la production de contenu académique et des tâches hybrides combinant ces deux fonctions. Ces observations corroborent les travaux de Jishnu et al. (2023) et de Goudey et al. (2024), qui soulignent le rôle croissant des IAG dans l'amélioration de la productivité académique, le soutien à l'apprentissage autonome et la simplification des processus cognitifs, notamment en termes de recherche d'information et de rédaction assistée.

Cependant, nos résultats révèlent également qu'environ 29,05 % des étudiants interrogés n'ont jamais eu recours à cet outil, et que 42,58 % des utilisateurs déclarés l'utilisent uniquement de manière occasionnelle. Cette situation met en lumière des barrières à l'intégration de ChatGPT en milieu éducatif et soulève des questions sur les disparités d'accès et de compétences dans l'usage de ces outils. Les non-utilisateurs et utilisateurs limités risquent d'être exclus des bénéfices pédagogiques offerts par l'IAG, comme le gain de temps ou l'optimisation des apprentissages (Madden et al., 2024 ; Lobet et al., 2023). Par ailleurs, ces résultats pourraient refléter une possible fracture numérique, où l'écart ne se limite pas à l'accès aux technologies, mais inclut également

les compétences nécessaires pour les utiliser efficacement, comme la formulation de prompts ou l'évaluation critique des réponses (Dai et al., 2023).

Dans notre étude, les utilisateurs limités se distinguent par des perceptions plus faibles de l'utilité et de l'utilisabilité de ChatGPT, ce qui reflète les conclusions de Madden et al. (2024) et du Pôle Léonard de Vinci (2024) sur l'importance de la formation pour exploiter pleinement le potentiel des outils d'IAG. Ces résultats soulignent également les risques liés à la vulnérabilité numérique, un enjeu clé selon Bachy (2024). Lorsqu'un outil d'IA est perçu comme facile à utiliser, les utilisateurs en tirent davantage de bénéfices et sont plus susceptibles de l'intégrer dans leurs pratiques régulières. Par ailleurs, Ahn (2024) montre que la confiance dans sa capacité à apprendre avec l'IAG favorise des améliorations significatives des résultats personnels et conduit à une utilisation accrue de ces outils. Ces observations indiquent que les utilisateurs limités pourraient rencontrer des obstacles à l'intégration de l'IA générative (IAG) dans leurs pratiques académiques, en raison d'un manque de compétences techniques ou d'une perception limitée de sa valeur ajoutée.

Pour ces étudiants, l'écart avec les utilisateurs réguliers est particulièrement préoccupant, car ces derniers développent des compétences techniques et cognitives spécifiques. En effet, on peut supposer, au vu de leur utilisation plus fréquente que le profil des utilisateurs hybrides, que ceux-ci tendent vers une maîtrise et une autonomie accrue, combinant recherche d'informations et production de contenu. Ce comportement s'aligne avec les compétences décrites dans le cadre de la littératie en IA (Preiksaitis & Rose, 2023), telles que la formulation efficace de prompts et l'évaluation critique des réponses générées. De plus, l'analyse des non-utilisateurs montre que ces derniers sont majoritairement plus jeunes avec une utilisation moins fréquente des outils numériques que les utilisateurs hybrides. Cette observation rejoint les travaux de Hurd & Orban de Xivry (2024), qui pointent les écarts d'adoption technologique selon l'âge et le niveau d'études. Il devient donc crucial de proposer des approches pédagogiques adaptées, particulièrement pour les élèves du secondaire, afin de réduire les inégalités d'accès et d'appropriation.

D'un point de vue pédagogique, cette recherche montre que l'adoption de ChatGPT n'est pas uniforme et dépend de variables complexes, telles que la fréquence d'utilisation des outils numériques, le niveau d'enseignement et la perception de ChatGPT. Les enseignants pourraient tirer parti de ces profils pour adapter leurs stratégies pédagogiques, en proposant des activités différenciées. Les profils identifiés montrent que l'IAG est perçue à la fois comme un facilitateur et une source d'incertitude, ce qui renforce la pertinence d'une littératie en IA pour garantir son intégration responsable (Preiksaitis & Rose, 2023). De ce fait, nous pouvons proposer les pistes pédagogiques suivantes. Les chercheurs d'informations utilisent ChatGPT principalement pour recueillir des données et valider des informations, reflétant un besoin de simplification des processus de recherche documentaire (Lobet et al., 2023). Ce profil met en évidence l'importance de former les étudiants à l'évaluation critique des réponses générées par l'IAG, en cohérence avec les recommandations de Dai et al. (2023) sur la validation des données. Une piste pédagogique pertinente consisterait à intégrer des activités basées sur la méthode de lecture latérale (Pavlounis et al., 2021),

où les étudiants comparent les informations produites par ChatGPT avec des sources académiques fiables, renforçant ainsi leur esprit critique et leur capacité à distinguer des données fiables. Les créateurs de contenu, quant à eux, exploitent ChatGPT principalement pour la production académique assistée. Inspiré des travaux de Cotton et al. (2023), un programme de formation pourrait inclure des ateliers sur la rédaction collaborative entre humains et IAG, soulignant la complémentarité entre créativité humaine et automatisation. Ces ateliers pourraient également s'appuyer sur les recommandations d'Ahajjam et Bouabbouz (2023) pour enseigner l'optimisation des prompts, une compétence clé pour exploiter efficacement le potentiel des IAG dans la production écrite.

Les utilisateurs hybrides utilisent ChatGPT pour des tâches variées, combinant recherche d'informations et création de contenu. Bien qu'ils démontrent une maîtrise avancée des outils d'IAG, ils pourraient bénéficier d'un approfondissement de leurs compétences en littératie numérique pour mieux intégrer l'IAG dans des contextes interdisciplinaires. Des formations avancées pourraient être proposées, telles que des projets collaboratifs où ChatGPT est utilisé comme assistant dans des tâches complexes, comme l'analyse de données ou la production de rapports multidimensionnels. Ces initiatives viseraient à consolider leurs compétences tout en les exposant à des applications plus spécialisées et créatives de l'IAG. Les utilisateurs limités, représentant le groupe le plus important, se distinguent par une utilisation sporadique et limitée de ChatGPT. Selon Lozano et Blanco Fontao (2023), ce comportement peut être lié à un manque de confiance dans les outils numériques ou à une méconnaissance de leurs fonctionnalités. Pour répondre à leurs besoins, des tutoriels interactifs ou des sessions de découverte en classe pourraient être organisés. Ces activités permettraient de les familiariser avec les fonctionnalités de base de ChatGPT dans un environnement encadré, réduisant ainsi leurs appréhensions et encourageant une utilisation plus régulière et bénéfique. Enfin, les non-utilisateurs, représentant 29 % des répondants, n'ont jamais utilisé ChatGPT. Ce comportement pourrait s'expliquer par des barrières spécifiques, telles qu'un accès limité à la plateforme ou une perception d'un manque de pertinence ou d'utilité dans leur contexte. Inspiré des travaux de Dai et al. (2023), une piste pédagogique pourrait consister à proposer des ateliers d'introduction similaires à ceux destinés aux utilisateurs limités. Ces ateliers offriraient une découverte guidée des fonctionnalités de ChatGPT et viseraient à sensibiliser les étudiants à l'intérêt de l'IAG pour leur apprentissage.

Cependant, la non-utilisation de ChatGPT pourraient ne pas résulter uniquement de barrières techniques, d'un manque de compétences ou d'une perception limitée de son utilité. Certains étudiants peuvent faire le choix éclairé de ne pas recourir à l'IAG pour des raisons éthiques, écologiques ou de confidentialité des données (Sano-Franchini et al., 2024 ; Ren & Wierman, 2024). Il convient de noter que les préoccupations liées à la confidentialité des données constituent un aspect spécifique des enjeux éthiques plus larges, mais que d'autres motifs, tels que la reproduction d'inégalités sociales ou l'exploitation de travailleurs précaires dans l'entraînement des algorithmes (Tubaro et al., 2024), peuvent également conduire à une décision de non-utilisation. Ce refus volontaire d'usage représente une posture critique légitime. Ces observations invitent à

élargir l'analyse des non-utilisateurs, en considérant non seulement les obstacles techniques, mais aussi les choix critiques et éthiques qui motivent une non-adoption consciente de l'IAG.

6 Conclusion

Cette étude propose une contribution à la compréhension des usages de ChatGPT en contexte éducatif, en identifiant des profils d'utilisateurs distincts construits à partir des dimensions fonctionnelles d'usage (informationnelle et productive) et des perceptions de l'outil. En combinant analyse d'usages et perceptions dans une population couvrant le secondaire et le supérieur, elle offre un éclairage sur la diversité des pratiques et sur les risques d'inégalités d'appropriation de l'IAG dans un contexte francophone encore peu documenté. Les résultats enrichissent ainsi le débat sur l'intégration des intelligences artificielles génératives dans l'éducation, en soulignant l'importance de former à une utilisation critique, éthique et écologiquement responsable.

Sur le plan pratique, les résultats plaident pour un accompagnement pédagogique différencié, visant notamment à soutenir les non-utilisateurs et les utilisateurs limités, afin de réduire les fractures d'usage et de maximiser les bénéfices éducatifs. Ils invitent également à reconsidérer les politiques éducatives, en intégrant l'IAG dans les curriculums tout en développant simultanément des compétences techniques, critiques et éthiques.

Cependant, plusieurs limites doivent être prises en compte. Le recours à un questionnaire auto-administré expose à des biais de déclaration (désirabilité sociale, imprécisions), ce qui peut affecter la validité des résultats. L'échantillon, limité à la Fédération Wallonie-Bruxelles et non différencié selon les types de formation ou les milieux socio-économiques, restreint la généralisation des observations. De plus, l'absence de données qualitatives ne permet pas d'explorer en profondeur les motivations, freins et positionnements éthiques vis-à-vis de l'usage ou du refus de l'IAG.

De futures recherches devraient combiner approches quantitatives et qualitatives pour analyser plus finement les dynamiques d'appropriation de l'IAG, en tenant compte du sentiment de tricherie, des enjeux de confidentialité des données, de l'impact écologique et des risques de dépendance technologique. Elles pourraient aussi interroger la légitimité d'une décision consciente de non-utilisation de l'IAG, fondée sur des considérations éthiques ou environnementales, et réfléchir à la manière de valoriser cette posture critique dans les dispositifs éducatifs. Enfin, l'expérimentation en contexte réel des dispositifs pédagogiques proposés, ainsi que l'étude de l'impact des biais algorithmiques sur les apprentissages, constituent des axes essentiels pour promouvoir une intégration éclairée et responsable de l'intelligence artificielle générative en éducation.

References

1. Abbasi, B.N., Wu, Y. & Luo, Z. (2024). Exploring the impact of artificial intelligence on curriculum development in global higher education institutions. *Education and Information Technologies*. https://doi.org/10.1007/s10639-024-13113-z
2. Ahajjam, A., & Bouabbouz, I. (2023, 9 mars). *Travailler sur ChatGPT sur une situation d'apprentissage.* [Communication]. Démystifier ChatGPT dans le domaine de l'éducation, Namur/Tournai.
3. Ahn, H. (2024). AI-Powered E-Learning for Lifelong Learners: Impact on Performance and Knowledge Application. *Sustainability*. https://doi.org/10.3390/su16209066.
4. An Ngo, T. (2023). The Perception by University Students of the Use of ChatGPT in Education. *International Journal of Emerging Technologies in Learning*, 17, 4-19. https://doi.org/10.3991/ijet.v18i17.39019
5. Bachy, S. (2024). Vulnérabilité numérique : un enjeu pour l'aide à la réussite. *Revue internationale des technologies en pédagogie universitaire*, *21*(1), 1-24. https://doi.org/10.18162/ritpu-2024-v21n1-01
6. Bitzenbauer, P. (2023). ChatGPT in physics education : A pilot study in easy-to-implement activities. *Contemporary Educational Technology, 15*(3). http://dx.doi.org/10.30935/cedtech/13176
7. Cabana, M & Lez, A. (2023). *Enseigner avec l'intelligence artificielle : Alternatives pour évaluer les apprentissages et stratégies visant l'optimisation de l'enseignement et de l'apprentissage (version 2).* Service de soutien à la formation, Université de Sherbrooke.
8. CLEMI. (2024). *Découvrir l'intelligence artificielle comme outil de recherche d'information.* https://www.clemi.fr/ressources/ressources-pedagogiques/decouvrir-lintelligence-artificielle-comme-outil-de-recherche-dinformation
9. Cotton, D., Cotton, P., & Shipway, R. (2023). Chatting and Cheating. Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*. https://doi.org/10.1080/14703297.2023.2190148
10. Degeer, M., & Kumps, A. (2022). *Les compétences numériques des élèves et des enseignants. A l'heure du Pacte pour un Enseignement d'excellence*. De Boeck.
11. Diallo, MF. (2023). Ce que ChatGPT fait à l'enseignement, à la recherche et aux organisations. *Revue Française de Gestion, 312*(5), 9-14. https://doi.org/10.3166/rfg.312.09-14
12. Goudey, A., Loupiac, P., & Quinio, B. (2024, 2 juillet). *L'impact des IA génératives dans l'enseignement supérieur*. SIDE Blog. https://alain.goudey.eu/side/2024/07/02/impact-ia-generatives-enseignement-superieur-gestion/
13. Fengchun, M. & Wayne, H. (2024). Orientations pour l'intelligence artificielle générative dans l'éducation et la recherche. *Organisation des Nations Unies pour l'éducation, la science et la culture [UNESCO].* https://doi.org/10.54675/HBCX3851
14. Floridi, L., & Chiriatti, M. GPT-3: Its Nature, Scope, Limits, and Consequences. *Minds & Machines, 30*, 681–694 (2020). https://doi.org/10.1007/s11023-020-09548-1
15. Hurd, S. & Orban de Xivry, A-C. (2024). *#Génération2024 : les jeunes et les pratiques numériques*.https://www.csem.be/sites/default/files/2024-05/Publication%20re%CC%81sultats-GEN2024.pdf
16. Jishnu, D., Srinivasan, M., Dhanunjay, G., & Shamala, R. (2023). Unveiling student motivations : a study of chatgpt usage in education. *ShodhKosh : Journal of Visual and Performing Arts, 4*(2), 65-73. https://dx.doi.org/10.29121/shodhkosh.v4.i2.2023.503

17. Li, J., Cheng, X., Zhao, X., Nie, J., & Wen, J-R. (2023). HaluEval: A Large-Scale Hallucination Evaluation Benchmark for Large Language Models. *In Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 6449–6464, Singapore. Association for Computational Linguistics.
18. Madden, M., Calvin, A., Hasse, A., & Lenhart, A. (2024). The dawn of the AI era: Teens, parents, and the adoption of generative AI at home and school. San Francisco, CA: Common Sense.
19. Lelong, P. (2023, 24 mars). *Chat GPT : Perspectives d'évolution de compétences et modes d'évaluation des enseignants.* [Communication]. L'intelligence artificielle et les compétences de demain, Charleroi.
20. Li, J., Cheng, X., Zhao, W.X., Nie, J., & Wen, J. (2023). HaluEval: A Large-Scale Hallucination Evaluation Benchmark for Large Language Models. *ArXiv, abs/2305.11747*.
21. Lobet, M., Honet, A., & Wathelet, V. (2023). *ChatGPT, l'usage par les étudiants de première année à l'université ChatGPT*. hal.science. https://shs.hal.science/halshs-04190226v1
22. Long, D., & Magerko, B. (2020). "What is AI literacy? Competencies and design considerations." Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems.
23. Lozano, A., & Blanco Fontao, C. (2023). Is the Education System Prepared for the Irruption of Artificial Intelligence? A study on the Perceptions of Students of Primary Education Degree from a Dual Perspective: Current Pupils and Future Teachers. *Education sciences, 13.* https://doi.org/10.3390/educsci13070733
24. Pavlounis, D., Johnston, J., Brodsky, J., et Brooks, P. (2021). L'urgence de repenser l'enseignement de la littératie numérique : Comment développer les aptitudes nécessaires pour faire face aux informations fausses et trompeuses en ligne, au moyen d'outils pédagogiques ayant fait leurs preuves. CIVIX Canada.
25. Pôle Léonard De Vinci. (2024, 16 mai). *L'impact des IA génératives sur les étudiants.* https://www.devinci.fr/le-pole-leonard-de-vinci-et-talan-publient-une-etude-edifiante-autour-de-limpact-des-ia-generatives-sur-les-etudiants/
26. Preiksaitis, C., & Rose, C. (2023). Opportunities, Challenges, and Future Directions of Generative Artificial Intelligence in Medical Education: Scoping Review. *JMIR Medical Education*, 9. https://doi.org/10.2196/48785.
27. Ren, S., & Wierman, A. (2024). The uneven distribution of AI's environmental impacts. *Harvard Business Review*. https://hbr.org/2024/07/the-uneven-distribution-of-ais-environmental-impacts
28. Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT : Bullshit spewer or the end of traditional assessments in higher education ? *Journal of Applied Learning & Teaching, 6*(1). https://doi.org/10.37074/jalt.2023.6.1.9
29. Sano-Franchini, J., McIntyre, M., & Fernandes, M. (2024). *Refusing Generative AI in Writing Studies: A Quickstart Guide***.** https://refusinggenai.wordpress.com
30. STEMENTIEL. (2024). *L'intelligence artificielle générative.* https://www.stementiel.be/ia-generative/
31. Sullivan, M., Kelly, A., & Mclaughlan, P. (2023). ChatGPT in higher education: Considerations for academic integrity and student learning. *Journal of Applied Learning & Teaching,6*(1). https://doi.org/10.37074/jalt.2023.6.1.17
32. Tubaro, P., Casilli, A. A., Fernández Massi, M., Longo, J., Torres Cierpe, J., & Viana Braz, M. (2025). The digital labour of artificial intelligence in Latin America: a comparison of Argentina, Brazil, and Venezuela. *Globalizations*, 1–16. https://doi.org/10.1080/14747731.2025.2465171

Analyse de l'engagement de l'étudiant lié à la difficulté d'items d'évaluation : Une étude d'EdNet fondée sur l'apprentissage automatique

Mohamed Lamgarraj[1][0009−0007−4708−1740], Céline Joiron[1][0009−0008−5782−8024], Aymeric Parant[2][0000−0003−0993−0661], and Gilles Dequen[1][0000−0001−7497−1182]

[1] Laboratoire de Modélisation, Information et Systèmes (MIS), Picardie Jules Verne university, Amiens 80000, France

[2] Centre de Recherche en Psychologie : Cognition, Psychisme et Organisations (CRP-CPO), Picardie Jules Verne university, Amiens 80000, France

Résumé . L'intégration des plateformes numériques dans l'éducation offre des opportunités pour suivre les activités des étudiants et utiliser ces données dans l'apprentissage automatique afin d'analyser les schémas d'apprentissage et leur impact sur les performances des étudiants. Cette étude s'inscrit dans le cadre de nos travaux de recherche sur l'analyse et l'exploitation des traces pédagogiques pour optimiser les systèmes de formation numérique. S'appuyant sur le jeu de données EdNet, comprenant plus de 100 millions d'interactions issues d'une plateforme d'apprentissage et d'évaluation numérique, nous examinons les corrélations entre les activités des apprenants, les contenus consultés, leurs taux de réussite, et la difficulté des questions proposées. En poursuivant nos travaux antérieurs sur la caractérisation de la difficulté des questions, nous appliquons des modèles d'apprentissage automatique pour prédire la réussite des étudiants, et identifier les principaux facteurs influençant cette réussite, tels que la nature des interactions, les niveaux de difficulté, et l'exposition aux contenus pédagogiques.

Mots-clés. Traces pédagogiques, Évaluation des Apprentissages, Difficulté, Apprentissage automatique, Intelligence artificielle (IA).

Abstract. The integration of digital platforms in education offers opportunities to track student activities and leverage this data in machine learning to analyze learning patterns and their impact on student performance. This study is part of our research on the analysis and use of educational traces to optimize digital training systems. Using EdNet dataset, which includes over 100 million interactions from a digital learning and assessment platform, we examine correlations between learners' activities, the content they engage with, their success rates, and the difficulty of the questions presented. Building on our previous work on characterizing question difficulty, we apply machine learning models to predict student success. These models also help identify the key factors influencing this success, such as the nature of interactions, levels of difficulty, and exposure to educational content.

Keywords: Learning Analytics · Data Mining · Student Performance · Question Difficulty · Artificial Intelligence (AI)

1 Introduction

Nos travaux s'inscrivent dans la cadre d'une recherche sur l'évaluation des apprentissages dans les systèmes numériques, et plus spécifiquement l'apport des techniques d'apprentissage automatiques dans la compréhension et la proposition d'évaluations et d'items d'évaluation dont on peut estimer/prédire la difficulté. En effet, avec l'évolution rapide des technologies et l'essor des plateformes numériques dans le domaine de l'éducation, la collecte et l'analyse des données éducatives offrent des opportunités pour mieux comprendre les facteurs influençant la réussite académique des étudiants, en particulier lors des test d'évaluation en ligne. De nombreuses études se sont concentrées sur les effets de l'engagement envers le contenu (l'ensemble des interactions de l'apprenant avec les ressources pédagogiques, telles que la consultation de contenu explicatifs, les ressources lié au sujet, ou encore le temps passé sur ces éléments.), les niveaux de difficulté des exercices, et leur impact sur les performances.

Les recherches montrent que l'interaction avec des supports éducatifs riches et adaptés peut améliorer significativement la compréhension et les résultats des étudiants. Des travaux tels que ceux de Chi et Wylie, 2014 ont introduit le concept de "tutoring cognitif actif", soulignant que l'engagement actif, mesuré par des interactions fréquentes et prolongées avec les contenus, favorise une meilleure mémorisation et une compréhension approfondie [4]. Pour le Séquençage des contenus, Des études ont également montré que l'ordre dans lequel les étudiants accèdent au contenu (par exemple, regarder des explications avant de répondre à des questions) a un impact direct sur leur réussite [13].

D'un autre côté, la difficulté des questions joue un rôle crucial dans le maintien de l'équilibre entre motivation et apprentissage. Par exemple la théorie de Zone proximale de développement (ZPD), inspirée des travaux de Vygotsky, suggère que les questions doivent être suffisamment difficiles pour stimuler l'apprenant, mais pas au point de décourager ses efforts [11]. Cette approche a été explorée dans des systèmes adaptatifs tels que ASSISTments [9]. Dans la modélisation des performances, des algorithmes tels que Knowledge Tracing (KT) et ses variantes améliorées (Bayesian KT, Deep KT) ont permis de modéliser la probabilité qu'un étudiant réponde correctement à une question donnée en fonction de la difficulté et de son historique d'apprentissage [16] .

Peu de recherches croisent directement l'impact de l'engagement, avec celui de la difficulté des questions, sur les performances globales des étudiants. Nous pouvons mentionner des travaux sur l'interaction entre engagement et adaptativité, qui ont exploré comment les plate-formes adaptatives, telles que ALEKS et Duolingo, modifient dynamiquement la difficulté des questions en fonction du niveau d'engagement de l'utilisateur, obtenant ainsi des gains d'apprentissage significatifs [7]. Enfin, une étude sur les plate-formes EdTech a montré que les

explications interactives, lorsqu'elles sont combinées avec des questions adaptées, augmentent les performances des étudiants de 25 % en moyenne [15].

Bien que ces études aient été menées sur chacun des aspects importants dans le knowledge-tracking, peu de travaux relient directement l'engagement, la difficulté des questions, et le succès des étudiants dans un cadre unifié. Cependant, une analyse intégrée nous semble nécessaire pour comprendre les interactions complexes entre engagement, difficulté des questions et succès des étudiants. Cela pourrait servir de base pour concevoir des systèmes encore plus performants et adaptatifs à l'avenir.

L'objectif principal de nos travaux est d'explorer la relation entre l'engagement des étudiants avec le contenu éducatif, les niveaux de difficulté des questions auxquelles ils sont confrontés, et leur succès académique. Ils visent à analyser comment les étudiants interagissent avec différents types de contenu — que ce soit des questions, des leçons ou des explications — et comment ces interactions influencent leur réussite. Aussi la manière dont la difficulté des questions, en relation avec l'engagement des étudiants, impacte leur capacité à réussir les tests. Cette étude exploite les données issues des interactions des étudiants avec le contenu éducatif sur une plate-forme numérique d'apprentissage, et disponibles dans la littérature (le dataset Ednet).

Cet article présente en premier lieu la constitution d'un jeu de données riche et diversifié en exploitant les traces d'utilisation disponibles dans Ednet [5], et en cherchant à obtenir un maximum d'informations pertinentes concernant les questions, les utilisateurs et le contexte d'apprentissage. Cette approche nous permet de bâtir une base de données capable de refléter les interactions complexes entre les étudiants et le contenu éducatif. Une fois le jeu de données constitué, nous détaillerons la méthodologie adoptée, qui repose sur l'application de techniques de data mining pour analyser les données. Ces méthodes nous permettent d'analyser les comportements des étudiants et d'examiner l'impact de leur engagement sur leur réussite académique. Enfin, nous discuterons des résultats obtenus et de leurs implications pratiques, notamment dans le cadre de l'évaluation numérique des apprentissages. .

2 Les données

Pour étudier la relation entre la consultation des ressources pédagogiques et les taux de réussite des apprenants, il est essentiel de disposer d'un ensemble de données riche et diversifié qui dépasse les indicateurs d'engagement traditionnels. Une compréhension approfondie de cette relation nécessite des données qui capturent non seulement les résultats des tests mais aussi l'accès aux ressources. Enfin, l'ensemble de données doit être suffisamment vaste et diversifié pour garantir que les modèles formés sur celui-ci généralisent bien sur divers sujets, thèmes et niveaux de difficulté. Les données que nous préparerons doivent alors être d'une échelle importante, répondre au critère de variété, pour pouvoir identifier des variables nuancées, et aux critères de qualité, c'est-à-dire d'authenticité, de cohérence et de granularité afin de garantir l'efficacité de modèles prédictifs.

Les données utilisées dans cette étude sont les données d'EdNet [5], une des plus grandes bases des traces pédagogiques ouvertes . Elle contient plus de 100 millions d'interactions étudiant-contenu, et inclut des informations détaillées sur les performances des étudiants. EdNet [5] se présente comme un ensemble de données hiérarchique, rassemblant deux ans de journaux d'interaction des étudiants provenant de Santa [1], une solution d'auto-apprentissage multi-plateforme conçue pour aider les étudiants à se préparer au test TOEIC (Test d'anglais pour la communication internationale). En plus des tests d'évaluation, les étudiants ont accès à une variété de ressources pédagogiques de formation en langue anglaise (vidéos, lectures, explications), permettant de renforcer leur compréhension des concepts et de se préparer de manière optimale. Le jeu de données EdNet est composé de 131 441 538 interactions collectées auprès de 784 309 utilisateurs. Elle comprend 19,4 Go de données réparties en 1,6 million de fichiers individuels au format CSV, structurés en quatre niveaux (KT1 à KT4), chacun fournissant des informations de plus en plus détaillées sur les actions des étudiants. KT1 se concentre sur les réponses aux questions, KT2 inclut aussi les interactions avec les explications et les exercices, KT3 couvre la consommation de vidéos et de supports pédagogiques, et KT4 ajoute des éléments sur la gestion du temps et l'ensemble des actions des étudiants sur la plate-forme. Les interactions des étudiants englobent leurs actions avec le matériel pédagogique (questions, vidéos, explications), leurs réponses et le temps passé sur chaque activité. Ces données permettent d'analyser leurs stratégies d'apprentissage et leur gestion du temps. Les questions et bundles représentent des ensembles de problèmes et de cours (plus de 13 000 problèmes et 1 000 cours), regroupés par thème, passage ou média. Les étudiants doivent résoudre toutes les questions d'un bundle pour le compléter, ce qui permet de suivre leur progression dans un contexte cohérent d'apprentissage. Enfin un fichier nommé "Contents", contient des informations sur les explications et les cours, et intègre les actions liées à la consultation d'explications (commentaires d'experts) et à la visualisation de vidéos. Chaque question ou item d'évaluation est associé aux contenus à l'aide d'étiquettes (tags). Ces supports sont cruciaux pour étudier l'impact de consultation de contenue sur la performance des étudiants et leur compréhension. En résumé, EdNet offre une vue d'ensemble détaillée des actions des étudiants à travers plusieurs niveaux, en analysant leur engagement avec différents types de contenu et leur gestion du temps d'apprentissage.

2.1 Présentation, enrichissement et préparation des données EdNet

Afin de constituer un ensemble de données pertinent, répondant à nos besoins de recherche, nous avons mené un processus rigoureux de regroupement, d'organisation et d'agrégation de l'ensemble des données EdNet pour nos premières tâches de recherche. L'ensemble EdNet, d'une taille de 19,4 Go et composé de 1,6 million de fichiers .csv, a servi de base pour cette analyse. Pour garantir la richesse des données et leur adéquation avec nos objectifs, Plusieurs étapes de préparation ont été effectuées lors de nos premières tâches de recherche, qui visaient à identifier les éléments influençant la difficulté perçue et susceptibles d'aider à prédire

celle-ci [14]. Ces étapes ont permis de structurer et d'optimiser les données de la façon suivante :

- Fusion des traces des apprenants : regroupement des 784,309 fichiers CSV relatifs à chaque utilisateur présents dans Ednet-KT1 en un fichier plat intégrant l'identifiant de l'utilisateur. Le résultat est un fichier composé de 95 293 926 lignes et 6 colonnes.

- Insertion des données relatives aux questions, et au contenu, issues du fichier 'Contents' telles que la réponse correcte à la question, le contenu associé à chaque question, les thèmes ou d'autres méta-données pertinentes.

- Amélioration des données : augmentation des données extraites avec des caractéristiques calculées et déduites. Ces caractéristiques conçues peuvent inclure des ratios, des statistiques, ou d'autres métriques dérivées des données initiales qui fournissent une compréhension plus approfondie des motifs sous-jacents dans l'ensemble de données.

Cette procédure élargit la portée de notre enquête tout en nous fournissant un ensemble de variables plus complet, susceptible de révéler des liens non découverts et d'aider à une meilleure prise de décision. Une série d'expérimentations progressives, consistant à ajouter ou retirer certaines caractéristiques afin d'identifier la combinaison produisant les meilleurs résultats en termes de corrélation avec la réussite des étudiants. Les informations et caractéristiques supplémentaires que nous avons intégrées au jeu de données sont comme suit :

Caractéristiques utilisateur :
mean_user_accuracy : taux de réussite moyen de chaque utilisateur.
question_answered : nombre total de questions répondues par un utilisateur.
AverageTimeperUser : temps moyen passé par l'utilisateur sur une question.
question_explanation : indique si l'étudiant a vu une explication.

Caractéristiques des questions :
answered_correctly : résultat de l'interaction (1 pour une bonne réponse, 0 pour une mauvaise réponse).
tags_community : groupes d'étiquettes associées aux questions, créées par des algorithmes de clustering présenté en section 3.2.
mean_content_accuracy : taux de réponse correcte pour chaque question.
difficulty_level : niveau de difficulté basé sur le taux de bonnes réponses (et notre modèle de prédiction).

Caractéristiques combinées :
performance_gap : Différence entre la précision moyenne d'un utilisateur et le taux de réussite moyen d'une question.
Les données finales sont constituées des éléments initiaux de base de données, plus les éléments suivants que nous avons ajoutés (dérivés des données brutes) :

Table 1: Noms de colonnes de l'ensemble de données

user_id	question_id	correct_answer	user_answer
is_correct	elapsed_time	bundle_cl	explanation_id
part	tags-cl	question_answered	mean_user_accuracy
AverageTimeperUser	AverageTimeforCorrect	averageTimeforFalse	performance_gap
question_explanation	difficulty (GT)	answered_correctly	tags_community

2.2 Exploration et analyse des données

À la suite des étapes de préparation et d'enrichissement, notre ensemble de données final regroupe les traces d'interaction de 393 656 utilisateurs uniques qui ont effectué 101 230 332 interactions. Ainsi, en moyenne, un utilisateur a effectué 257,15 interactions. Ces données vont nous permettre d'identifier des tendances entre certaines caractéristiques des interactions (comme la durée moyenne par réponse ou le taux d'engagement vis à vis des tags de contenus) et la probabilité de répondre correctement à une question ('answer_correctly'). Il convient alors de pré-traiter les données et de créer des caractéristiques permettant aux modèles de prédire la probabilité d'exactitude. Ainsi, la relation entre la variable et les autres caractéristiques doit être étudiée.
Dans cette phase, nous procédons à une Analyse des Données Exploratoire (EDA), qui a pour objectif d'examiner en détail les relations entre les différents éléments de l'ensemble de données. Cette étape permet non seulement d'identifier les connexions et dépendances potentielles entre les variables (comme la difficulté des questions, le temps de réponse, ou les communautés de tags), mais aussi d'évaluer la pertinence des données pour la tâche d'apprentissage menée. En explorant les interactions et les caractéristiques associées, nous cherchons à extraire des statistiques descriptives et des observations clés, qui guideront les étapes suivantes de modélisation.
Parmi les premières observations, nous notons que 65,73% des questions ont été correctement répondues par les utilisateurs (Fig. 1). Ce taux de réussite global reflète une performance relativement positive des apprenants. Il fournit également des indications sur l'équilibrage de notre base de données nécessaire pour analyser l'impact des caractéristiques des questions et des contenus éducatifs sur les résultats obtenus.

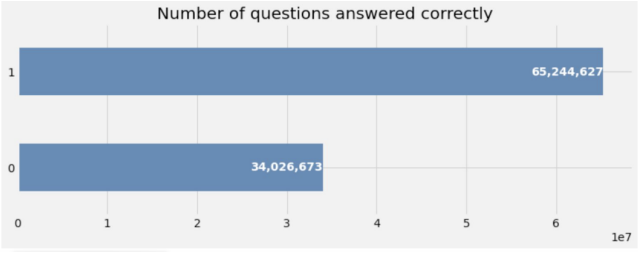


Fig. 1: Nombre des questions correctement répondues '1', avec des reponses fausses '0'

Chaque question peut avoir une ou plusieurs étiquettes associées, représentant des compétences spécifiques évaluées dans la question. Un premier travail consiste donc à encoder ces combinaisons d'étiquettes de façon à pouvoir les exploiter via l'analyse des données. Etant donné le caractère pédagogique des items, nous partons de l'hypothèse que certaines étiquettes sont regroupées plus fréquemment avec d'autres. Ainsi, plutôt que de stocker chaque tag associé à une question, l'idée est d'associer les informations relatives aux groupes de tags auxquels chaque item fait référence. Il s'agit alors de regrouper les tags en communautés en les divisant selon leurs clusters, et d'utiliser ces derniers pour encoder les étiquettes. Un premier graphe a été construit à l'aide du module Networkx. Chaque étiquette y est représentée par un nœud, et les arêtes indiquent les co-occurrences des étiquettes dans les items de la base. Le poids de chaque arête représente le nombre de questions dans lesquelles deux étiquettes adjacentes apparaissent ensemble(Fig. 2).

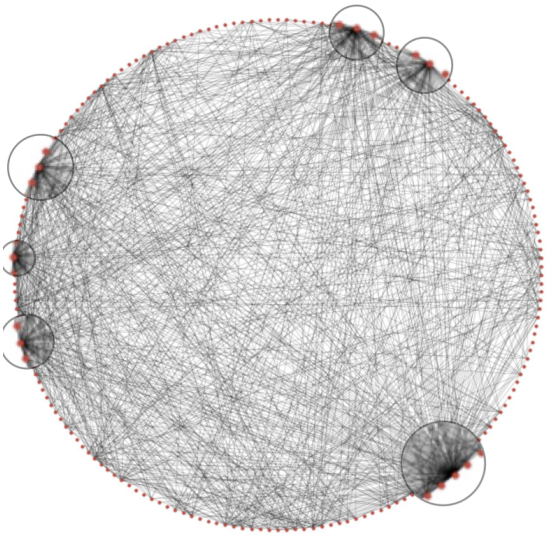


Fig. 2: Graphe des associations d'étiquettes d'items (tags)

Ce graphe révèle que certains nœuds forment des zones plus denses, ce qui indique que les étiquettes associées à ces nœuds apparaissent plus fréquemment que d'autres. Une analyse plus approfondie montre également que certains nœuds ne possèdent aucune arête ce qui indique que certaines étiquettes apparaissent toujours seules dans les items de la base. Ces dernières ont donc été isolées afin de poursuivre l'observation des motifs de regroupement en traçant à nouveau le graphe avec les étiquettes restantes(Fig. 3). Ce second graphe mets en évidence que certains nœuds ne sont connectés qu'au nœud central. Ainsi, ces étiquettes apparaissent soit seules, soit en paire avec l'étiquette représentée par le nœud central, que nous identifions comme étant l'étiquette 162. Ces étiquettes ayant une seule connexion sortante ont été listées, ainsi que le nombre de fois où elles

apparaissent en paire avec l'étiquette 162. Enfin, un troisième graphe permet de représenter les nœuds restants, qui présentent des associations intrinsèques plus complexes (Fig. 4). Ce dernier établit l'existence de deux regroupements. Un algorithme de détection de communautés permettant de séparer les nœuds en deux clusters distincts puis de les colorer en rouge et bleu confirme cette observation.

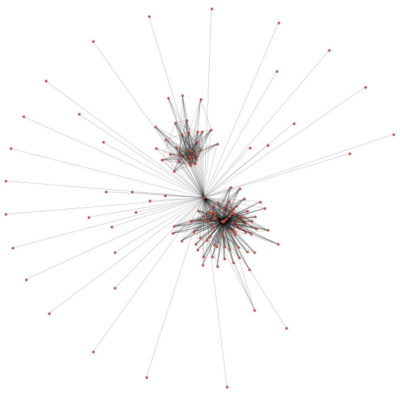


Fig. 3: Etiquettes par rapprochement des nœuds les plus liés

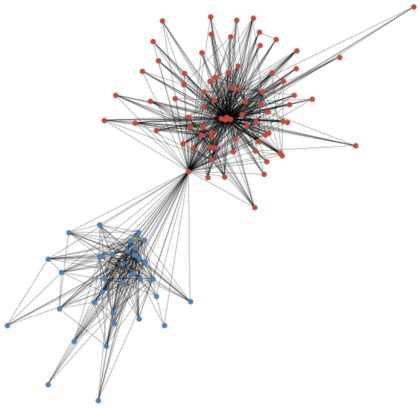


Fig. 4: Groupes d'étiquettes

Les étiquettes de 12 284 questions ont été catégorisées en quatre communautés distinctes : une liste contenant toutes les étiquettes qui apparaissent uniquement de manière isolée (communauté 0) ; une liste regroupant toutes les étiquettes qui apparaissent seules ou en paire avec l'étiquette 162 (communauté 1); deux clusters révélées à partir des étiquettes restantes (communautés 2 et 3). Le tableau (Tab. 2) présente la répartition des items par communauté.

Table 2: la répartition des questions sur les 4 groupes de tages

tags Community	0	1	2	3
number of questions	512 (4.17%)	2,045 (16.64%)	5,972 (48.62%)	3,755 (30.57%)

En plus des étiquettes thématiques pour chaque question, les données EdNet classifient les items en différentes parties, correspondant à des types spécifiques d'activités éducatives réalisées par les apprenants. Ces parties se décomposent en sections telles que Écoute comprend les parties 1 à 4 (incluant des activités comme Photographies, Questions-Réponses, Conversations et Discours) et Lecture dans les parties 5 à 7 (comprenant des activités comme Phrases Incomplètes, Complétion de Texte, et Passages Simples ou Multiples). Ces différentes parties reflètent la diversité des compétences sollicitées dans le processus d'apprentissage et dévaluation, et peuvent nous permettre d'étudier des patterns d'interactions spécifiques. Ainsi, la figure ci-dessous illustre la répartition des réponses correctes

et incorrectes par partie, offrant un aperçu des performances des apprenants en fonction du type d'activité (Fig. 5).

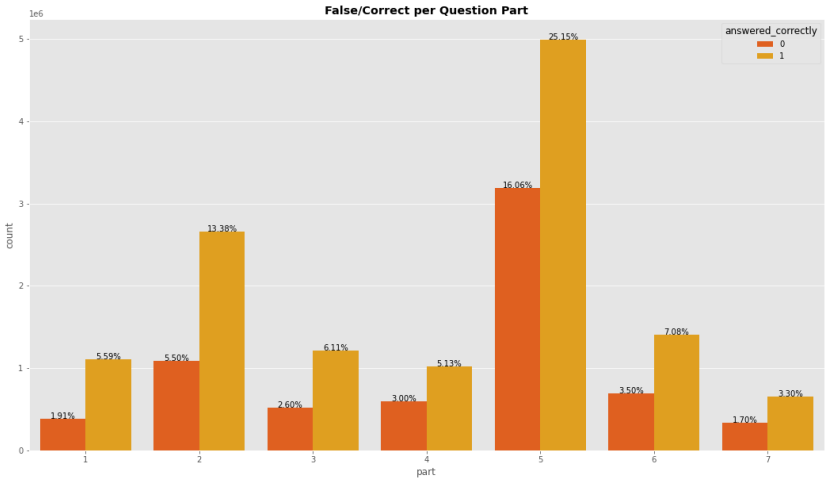


Fig. 5: Répartition des réponses correctes et incorrectes par parties

La section suivante explore les modèles expérimentés sur ces ensembles de données afin d'évaluer leur capacité à prédire la réussite des apprenants, au regard de la difficulté des questions, et à mettre en évidence les relations entre les différents contenus, thématiques ou activités pédagogiques des apprenants. Les résultats obtenus sont ensuite analysés.

3 Modèles et résultats

Après avoir identifié les données significatives pour les tâches d'apprentissage automatique, notre approche consiste à entraîner des modèles capables d'exploiter cet ensemble de données pour fournir des représentations précises de l'engagement de l'apprenant, tout en permettant une estimation robuste de son impact sur la réussite des apprenants aux items. Pour cet entraînement il convient en premier lieu de sélectionner un échantillon représentatif des données issues de la base de données construite comme préconisé dans [2]. Ainsi, Les profils utilisateurs sélectionnés couvrent une diversité de comportements, incluant des apprenants sans interaction et des utilisateurs modérément et fortement actifs, selon le nombre total d'interactions. Ces derniers représentent 40% des profils sélectionnés de façon à analyser avec précision les comportements d'apprentissage les plus significatifs. Les questions ayant un nombre de passages compris entre 10 et 100 utilisateurs ont été sélectionnées afin de garantir une représentativité équilibrée entre popularité et difficulté. De plus, seules les données des interactions enregistrées durant la même période ont été utilisées pour s'assurer de leur fraîcheur et pertinence. Enfin, nous nous sommes assurées que les utilisateurs sélectionnés

soient diversifiés, composés d'utilisateurs sans interactions avec le contenu et d'utilisateurs qui ont interagi avec des types de contenus différents à des fréquences variées, afin de capturer une diversité de comportements et d'optimiser la richesse des analyses.
Une série d'expériences permet d'évaluer l'importance des différents éléments de la base de données dans la prédiction de la variable cible indiquant la réussite (answered_correctly).Plusieurs modèles d'apprentissage automatique ont été utilisés afin de mesurer l'impact de chaque ajout progressif d'éléments sur les performances prédictives. Les modèles sélectionnés incluent LightGBM, XGBoost et un réseau de neurones implémenté avec Keras. Chaque modèle a été entraîné sur 50 % des données et testé sur l'autre moitié, en utilisant l'AUC (Area Under Curve) [8] comme métrique principale pour évaluer la performance (Tab. 3).Une répartition équivalente (50/50) entre les données d'entraînement et de test a été retenue afin de maximiser la représentativité de l'échantillon de test et d'éviter le sur-apprentissage, tout en conservant un volume suffisant pour entraîner des modèles performants. La métrique AUC (Area Under Curve) a été utilisée pour évaluer les performances des modèles de classification, car elle est particulièrement adaptée aux problèmes où les classes sont déséquilibrées [6], comme dans notre cas. L'AUC mesure la capacité d'un modèle à distinguer correctement entre deux classes (par exemple, réussite ou échec). Elle est calculée comme l'aire sous la courbe ROC (Receiver Operating Characteristic), qui représente graphiquement le compromis entre le taux de vrais positifs (True Positive Rate) et le taux de faux positifs (False Positive Rate) à différents seuils de classification.

L'entraînement des modèles a été réalisé sur une infrastructure de calcul dédiée fournie par l'université **MatriCS** [17], plus précisément la partition comprenant les environnements **bigmem** et **risk-bigmem**. Ces environnements sont composées de 12 serveurs bi-processeurs Intel Xeon E5-2680 v4 à 2,40 GHz, chacun équipé de 28 cœurs (14 par processeur) et prenant en charge jusqu'à 56 threads grâce à l'architecture Hyper-Threading. Chaque serveur dispose de 512 Go de mémoire vive pour traiter de grands ensembles de données, et utilise un système de stockage haute performance optimisé pour le traitement intensif des données.

Ces ressources ont permis d'entraîner efficacement les modèles sur des données volumineuses, tout en maintenant des temps d'exécution raisonnables pour les expérimentations et les validations croisées. Cette approche permet de mieux comprendre la contribution de chaque élément de la base de données dans la construction d'un modèle prédictif performant. Les résultats obtenus révèlent les différences de performance entre les modèles et mettent en lumière les relations clés entre les caractéristiques extraites et les prédictions finales.
Les modèles utilisés ainsi que les résultats optimaux obtenus sont les suivants :

LightGBM (Gradient Boosting) [12]: Modèle principal utilisé grâce à sa capacité à traiter de grands ensembles de données avec une faible mémoire. LightGBM est une implémentation des arbres de décision à gradient boost-

ing. Le boosting est une technique d'ensemble qui consiste à ajouter de nouveaux modèles successivement pour corriger les erreurs des modèles existants, jusqu'à ce qu'il ne soit plus possible d'améliorer les performances. De plus, le gradient boosting utilise l'algorithme de descente de gradient pour minimiser la perte lors de l'ajout de nouveaux modèles destinés à prédire les erreurs des modèles existants. Les prédictions finales sont ensuite obtenues en sommant l'ensemble des modèles construits.

Paramètres clés :

- `num_leaves` : Contrôle la complexité de l'arbre.
- `max_depth` : Limite la profondeur des arbres pour éviter le sur-apprentissage.
- `early_stopping_rounds` : Arrêt anticipé pour prévenir le sur-apprentissage.

Avantages : Résultat : AUC (Area Under Curve) de **0.760**.

XGBoost (eXtreme Gradient Boosting) [3]: Utilisé pour comparaison, avec des paramètres similaires à ceux de LightGBM. Résultat : AUC de **0.754**, légèrement inférieur à LightGBM.

Keras Neural Network [10]: Modèle de réseau de neurones basé sur des couches denses et convolutives. Utilisé pour comparer avec les modèles de boosting. Résultat : AUC de **0.746**, moins performant que LightGBM et XGBoost.

L'utilisation de l'AUC dans notre étude s'appuie sur ses propriétés adaptées aux caractéristiques de nos données et sur sa capacité à fournir une évaluation fiable des modèles appliqués. Les résultats montrent que le modèle LightGBM offre la meilleure performance en terme de discrimination, ce qui en fait le choix le plus pertinent pour explorer la relation entre les variables de notre base de données et l'objectif (le succès).

Table 3: Les résultas obtenus utilisant le métrique AUC .

Model	LightGBM	XGBoost	keras NN
AUC	0.760	0.754	0.746

D'après les résultats obtenus, les étudiants passant plus de temps sur les explications (en lectures précisément) ont montré une amélioration de 10% de leur précision par rapport à ceux qui ne les ont pas consultées(Fig. 6). La fréquence des interactions est positivement corrélée avec les performances globales, bien qu'une diminution des bénéfices ait été observée après un grand nombre d'interactions (500 interactions), suggérant des rendements décroissants au-delà d'un certain seuil d'engagement. Les étudiants qui interagissaient avec des explications avant d'aborder des questions difficiles ont montré une amélioration de 30 % de leurs performances, mettant en évidence l'importance d'une préparation adéquate face à des défis plus complexes.

L'analyse des taux de réussite en fonction des tags révèle des disparités significatives, indiquant l'existence de variations de difficulté entre les questions appartenant à des catégories similaires (Fig. 7). Ces résultats suggèrent une corrélation entre les types de contenus éducatifs et les performances des apprenants. Certaines questions, bien qu'associées à des classes identiques, présentent des taux de réussite nettement plus faibles, ce qui pourrait être attribué à des facteurs tels

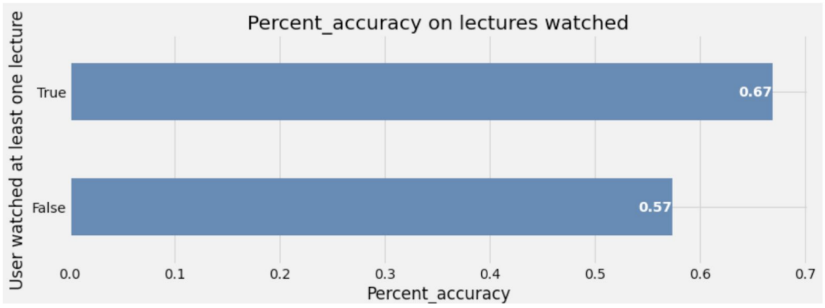


Fig. 6: Impact de l'engagement dans les lectures sur le taux de réussite

que la complexité contextuelle ou l'exigence cognitive. Cette observation met en évidence l'importance d'examiner non seulement le contenu éducatif lui-même, mais également la manière dont il est structuré et présenté, afin de mieux comprendre son influence sur les résultats académiques. Les figures suivantes mon-

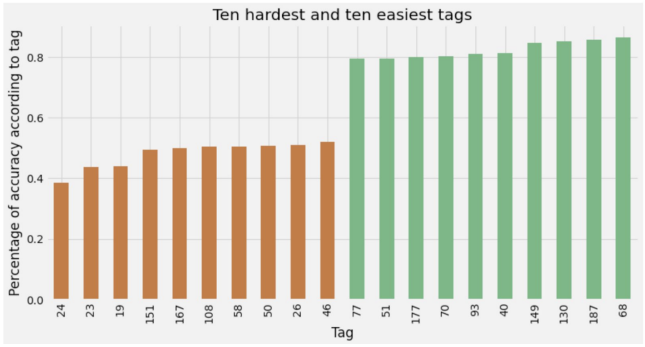
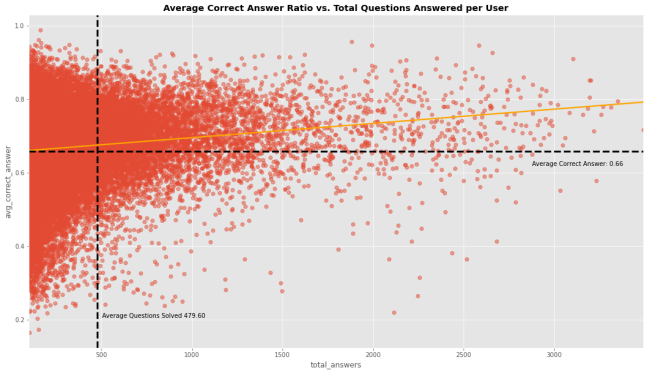


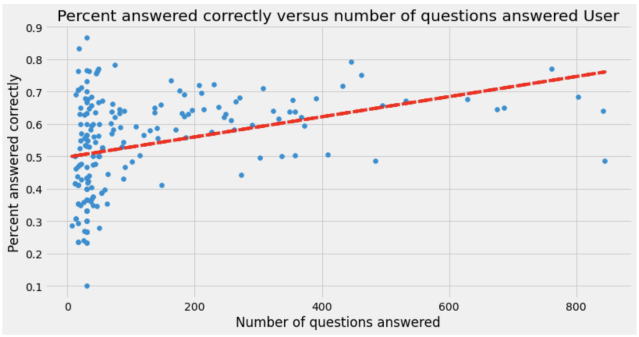
Fig. 7: Taux de réussite pour les 10 tags les plus difficiles et les 10 plus faciles

trent la distribution des réponses correctes en fonction du nombre total de questions répondues par utilisateur (Fig. 8), tant pour les données d'entraînement que pour les prédictions. Cette comparaison permet d'évaluer la performance du modèle en terme de cohérence entre les données réelles et les prédictions effectuées.

La cohérence observée entre les distributions des réponses correctes en fonction du nombre total de questions répondues, tant pour les données d'entraînement que pour les prédictions, suggère une bonne capacité de généralisation du modèle. En effet, les distributions similaires indiquent que les caractéristiques utilisées pour entraîner les modèles sont pertinentes et significatives pour prédire le succès des étudiants. Parmi ces caractéristiques, les colonnes relatives aux statistiques de visites des contenus, telles que la fréquence et le temps passé sur les explications, ainsi que la colonne de difficulté des questions, apparaissent comme des facteurs clés dans la performance prédictive.



(a) Distribution des réponses correctes en fonction du nombre total de questions répondues par utilisateur (données d'entraînement)



(b) Distribution des réponses correctes en fonction du nombre total de questions répondues par utilisateur (Prédictions)

Fig. 8: Comparaison de la distribution des réponses correctes en fonction du nombre total de questions répondues par utilisateur, entre les données d'entraînement et les prédictions.

Les statistiques de visite des contenus, en particulier, montrent une forte corrélation avec les réponses correctes, ce qui confirme l'importance de l'engagement actif des étudiants avec les ressources pédagogiques pour améliorer leur précision dans les réponses. De même, la difficulté des questions, bien que variée, s'est avérée un bon indicateur dans la capacité des étudiants à réussir les questions en fonction de leur niveau de préparation et de leurs interactions avec le contenu.

4 Conclusion

Cette étude met en évidence l'importance des systèmes de formation et d'évaluation numériques dans l'amélioration des performances académiques des étudiants. L'engagement des étudiants avec le contenu pédagogique et l'ajustement des niveaux de difficulté des questions sont des leviers essentiels pour favoriser la réussite. Nous avons identifié que l'utilisation régulière des explications pédagogiques,

telles que les feedbacks, améliore la compréhension et la précision des réponses des étudiants. De plus, l'adéquation entre la difficulté perçue des questions et les capacités des étudiants est cruciale pour optimiser l'apprentissage.
Notre travail s'inscrit dans une démarche visant à exploiter les données issues des traces pédagogiques pour identifier les facteurs influençant la réussite académique et la perception de la difficulté des éléments d'évaluation [14]. En continuité avec nos précédentes recherches sur la caractérisation et la prédiction des difficultés des questions. S'inscrivant dans ce démarche, cette étude a permis de mettre en lumière l'interaction entre divers facteurs clés des traces pédagogiques, notamment l'engagement des étudiants et leurs performances. Nous montrons que l'analyse des données pédagogiques peut concevoir des modèles prédictifs puissants, permettant d'adapter les contenus et évaluations aux besoins individuels des apprenants.
En dépassant les limites des approches théoriques traditionnelles, notre approche basée sur les données et l'apprentissage automatique, appliquée à un large volume de données d'interactions des apprenants sur une plate-forme numérique(EdNet), permet d'identifier des corrélations précises entre les activités des étudiants, la difficulté des questions et leur réussite. Nos analyses révèlent qu'une forte corrélation existe entre les facteurs clés extraits des bases de données et la réussite des étudiants. Par exemple, les étudiants passant plus de temps sur les contenus explicatifs ont un taux de réussite supérieur de 15-20% sur les questions complexes. Un séquencement optimal des tâches (simples avant complexes) améliore également les performances. Pour approfondir ces résultats, il est essentiel d'enrichir les bases de données avec des facteurs psychologiques et comportementaux, tels que la motivation, la gestion du stress et les styles d'apprentissage. Ces éléments, souvent sous-explorés, pourraient interagir de manière significative avec les facteurs académiques. En les intégrant aux traces d'interactions éducatives, nous pourrions non seulement améliorer les modèles prédictifs, mais aussi maximiser l'efficacité des systèmes d'apprentissage en ligne.

5 Remerciements

Ces travaux sont financés dans le cadre du projet ****** porté par ***** et résultant de l'appel à projet Démonstrateur Numérique de l'Enseignement Supérieur de France 2030.

References

1. Aitutor santa. https://www.aitutorsanta.com/vn/
2. Bengio, Y., LeCun, Y.: Scaling learning algorithms toward ai (2007)
3. Chen, T.: Xgboost: extreme gradient boosting. R package version 0.4-2 **1**(4) (2015)
4. Chi, M.T., Adams, J., Bogusch, E.B., Bruchok, C., Kang, S., Lancaster, M., Levy, R., Li, N., McEldoon, K.L., Stump, G.S., et al.: Translating the icap theory of cognitive engagement into practice. Cognitive science **42**(6), 1777–1832 (2018)

5. Choi, Y., Lee, Y., Shin, D., Cho, J., Park, S., Lee, S., Baek, J., Bae, C., Kim, B., Heo, J.: Ednet: A large-scale hierarchical dataset in education. In: Artificial Intelligence in Education: 21st International Conference, AIED 2020, Ifrane, Morocco, July 6–10, 2020, Proceedings, Part II 21. pp. 69–73. Springer (2020)
6. Davis, J., Goadrich, M.: The relationship between precision-recall and roc curves. In: Proceedings of the 23rd international conference on Machine learning. pp. 233–240 (2006)
7. Dong, L., Mallinson, J., Reddy, S., Lapata, M.: Learning to paraphrase for question answering. arXiv preprint arXiv:1708.06022 (2017)
8. Fawcett, T.: An introduction to roc analysis. Pattern recognition letters **27**(8), 861–874 (2006)
9. Feng, M., Heffernan, N., Koedinger, K.: Addressing the assessment challenge with an online system that tutors as it assesses. User modeling and user-adapted interaction **19**, 243–266 (2009)
10. Gulli, A., Pal, S.: Deep learning with Keras. Packt Publishing Ltd (2017)
11. Heffernan, N.T., Heffernan, C.L.: The assistments ecosystem: Building a platform that brings scientists and teachers together for minimally invasive research on human learning and teaching. International Journal of Artificial Intelligence in Education **24**, 470–497 (2014)
12. Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T.Y.: Lightgbm: A highly efficient gradient boosting decision tree. Advances in neural information processing systems **30** (2017)
13. Koedinger, K.R., Aleven, V.: Exploring the assistance dilemma in experiments with cognitive tutors. Educational Psychology Review **19**, 239–264 (2007)
14. Lamgarraj, M., Joiron, C., Parant, A., Dequen, G.: Exploring item difficulty prediction: Data driven approach for item difficulty estimation. In: International Conference on Intelligent Tutoring Systems. pp. 415–424. Springer (2024)
15. Liu, J., Yuan, H., Lu, X.M., Wang, X.: Quantum fisher information matrix and multiparameter estimation. Journal of Physics A: Mathematical and Theoretical **53**(2), 023001 (2020)
16. Piech, C., Bassen, J., Huang, J., Ganguli, S., Sahami, M., Guibas, L.J., Sohl-Dickstein, J.: Deep knowledge tracing. Advances in neural information processing systems **28** (2015)
17. Plateforme MatriCS: Plateforme matrics – université de picardie jules verne. https://www.matrics.u-picardie.fr/accueil/presentation-generale/ (2025), plateforme cofinancée par l'Union Européenne (FEDER) et le Conseil Régional des Hauts-de-France

Apprendre l'anatomie en dessinant sur tablette : effets du type d'activité et des capacités visuospatiales sur l'apprentissage et les perceptions d'étudiant·es en kinésithérapie

Laura Leconte[1], Eric Jamet[2]

[1] Univ Rennes, LP3C [(Laboratoire de Psychologie : Cognition, Comportement, Communication)], F-35000 Rennes, France
laura.leconte@univ-rennes2.fr
[2] Univ Rennes, LP3C [(Laboratoire de Psychologie : Cognition, Comportement, Communication)], F-35000 Rennes, France
eric.jamet@univ-rennes2.fr

Résumé. Cette étude a examiné l'efficacité de deux stratégies de dessin pour l'apprentissage de l'anatomie sur tablette-stylet : une activité de recopie et une activité de dessin réalisé de mémoire. 84 étudiant·es en deuxième année de formation en kinésithérapie ont appris un cours d'anatomie concernant les ligaments longitudinaux du rachis. Des analyses de régressions linéaires ont été menées afin d'étudier l'effet de la condition de dessin sur les performances d'apprentissage en fonction des capacités visuospatiales des étudiant·es. Les résultats ont montré que les processus cognitifs impliqués par l'activité de dessin réalisé de mémoire améliorent significativement le rappel. Les capacités visuospatiales se révèlent être déterminantes, favorisant la qualité des dessins produits et les performances d'apprentissage, indépendamment de la stratégie de dessin utilisée. Par ailleurs, la stratégie de dessin est globalement appréciée par les étudiant·es et perçue comme particulièrement utile pour faciliter l'apprentissage de l'anatomie.

Mots-clés : stratégie d'apprentissage par le dessin, apprentissage de l'anatomie, capacités visuospatiales, apprentissage sur tablette, apprentissage actif

Abstract. This study examined effectiveness of two drawing strategies for learning anatomy on a tablet with stylus: a copying activity and a drawing from memory activity. 84 students in second year of kinesitherapy were taught an anatomy course on longitudinal ligaments of the spine. Linear regression analyses were conducted to study effect of drawing condition on learning outcomes, according to students' visuospatial abilities. Results showed that cognitive processes involved in drawing from memory significantly improve recall. Visuospatial abilities were found to be determining factor, favoring quality of the drawing strategy used. In addition, drawing strategy was generally appreciated by students and seen as particularly useful in facilitating anatomy learning.

Keywords: learning by drawing strategy, anatomy learning, visuospatial abilities, tablet-based learning, active learning

1 Introduction

L'efficacité des stratégies d'apprentissage actives dans l'acquisition des connaissances est largement étudiée dans la littérature scientifique [1,2]. Parmi ces stratégies, celle qui consiste à faire réaliser des dessins se distingue par sa capacité à favoriser la construction de représentations visuelles des concepts et à améliorer la compréhension [3,4]. En transformant des contenus textuels en représentations imagées, cette activité générative engage des processus cognitifs profonds, tels que la sélection, l'organisation et l'intégration des informations [5,6]. Cependant, l'efficacité du dessin peut être influencée par la complexité du contenu et l'expertise des apprenant·es. De fait, une surcharge cognitive peut limiter les bénéfices de la stratégie de dessin, notamment chez les novices. Par ailleurs, les apprenant·es peuvent également se focaliser uniquement sur l'esthétique du dessin, oubliant l'objectif d'apprentissage [7–9].

Dans ce contexte, l'intégration des outils numériques offre des perspectives intéressantes pour enrichir l'expérience d'apprentissage par le dessin. Une méta-analyse récente [3] a montré que si les effets globaux de la stratégie de dessin sont bénéfiques, les résultats sont très hétérogènes et semblent influencés par de nombreux facteurs modérateurs. Parmi ces facteurs, le mode de dessin (dessin à la main vs dessin sur ordinateur) semble avoir une influence particulièrement forte. Alors que le dessin à la main s'est avéré avoir des effets positifs sur divers résultats d'apprentissage dans cette méta-analyse, aucun effet n'a été observé dans les études où le dessin a été effectué à l'aide d'un ordinateur. Selon les auteurs, ces résultats sont probablement dus à la complexité du dessin à l'aide d'une souris d'ordinateur car, dans presque toutes ces études, les dessins ont été réalisés à l'aide d'une méthode de « glisser-déposer ». Toutefois, il est possible aujourd'hui d'utiliser des dispositifs numériques plus adaptés à cette activité.

Les tablettes avec stylet, en particulier, offrent la possibilité de créer des contenus pédagogiques variés et interactifs. Plusieurs études ont mis en avant le rôle des tablettes dans l'amélioration des apprentissages, l'acquisition des connaissances [10,11], et la motivation des étudiant·es [12,13]. Cependant, malgré leur potentiel, l'impact des tablettes sur des stratégies spécifiques comme le dessin reste encore peu exploré [14–17].

C'est dans cette perspective que s'inscrit l'étude présentée ici. Elle est menée dans le cadre du projet ANR SKETCH dont les objectifs sont 1) d'évaluer et d'optimiser les effets des stratégies de dessin sur l'apprentissage, notamment en fournissant aux apprenant·es différentes formes de support pendant l'activité, et 2) de concevoir de manière collaborative un système tutoriel intelligent (STI) qui soit capable d'analyser de manière incrémentale les actions de l'apprenant·e lors de l'élaboration à main levée d'un croquis d'anatomie. L'usage de la tablette numérique avec stylet constitue un élément central du projet, permettant de reproduire l'activité de dessin traditionnellement réalisée sur papier-crayon, tout en offrant des possibilités d'environnement d'apprentissage adaptées et enrichies. Dans la présente étude, la segmentation de l'activité en plusieurs étapes, ainsi que la visualisation dynamique des tracés a permis aux apprenant·es de

suivre en temps réel la construction progressive du dessin et d'apprendre à leur propre rythme.

En lien avec le premier objectif du projet, le but de cette étude est d'évaluer l'efficacité de deux stratégies distinctes de dessin réalisées sur tablette dans le contexte de l'apprentissage de l'anatomie. La particularité de l'apprentissage de l'anatomie réside dans son exigence envers les apprenant·es de développer des connaissances spatiales. Comprendre et mémoriser les structures anatomiques nécessite une représentation mentale précise de leurs formes, leurs tailles et leurs apparences. L'activité de dessin semble donc appropriée pour l'apprentissage de l'anatomie. Plusieurs travaux dans ce domaine ont mis en avant les perceptions et ressentis des apprenant·es utilisant cette stratégie pendant l'apprentissage, et montrent que celle-ci est généralement bien accueillie [18–20]. Cependant, ces études ne mesurent pas directement l'impact de l'activité de dessin sur les performances d'apprentissage [21]. Par ailleurs, la stratégie d'apprentissage telle qu'étudiée en tant que stratégie générative, est réalisée à partir de la lecture d'un texte et non d'une illustration déjà fournie. Or, en l'absence de représentations visuelles, il devient difficile pour les étudiant·es d'intégrer correctement les connaissances anatomiques. Ainsi, l'importance du visuel dans la transmission des connaissances anatomiques rend presque obligatoire l'utilisation de représentations picturales (dessins ou croquis anatomiques).

Dans ce contexte, deux stratégies de génération de dessin d'apprentissage seront évaluées dans cette étude : la recopie d'un modèle et le dessin de mémoire. La première stratégie consiste à reproduire un modèle fourni d'une structure anatomique à apprendre. Cette méthode s'apparente à la stratégie de dessin décrite dans la littérature, à la différence que le dessin est élaboré à partir d'un document multimédia, contenant du texte ainsi qu'une illustration. Cependant, dans ce contexte multimédia, la stratégie de génération par recopie pourrait ne pas être aussi efficace. Bien qu'elle permette de produire des dessins d'apprentissage de bonne qualité, elle risque d'encourager un traitement cognitif relativement passif de recopie, limitant ainsi l'engagement actif recherché et, en conséquence, la qualité de la représentation en mémoire.

Afin d'encourager l'élaboration de cette représentation interne, il est possible d'éviter l'effet de recopie en faisant dessiner le modèle de mémoire. Dans cette condition, les apprenant·es doivent reproduire un dessin à partir de leur mémoire, après avoir été exposé au modèle. Cette approche repose sur une récupération active, qui mobiliserait les informations mémorisées et renforcerait leur consolidation. Bien qu'elle puisse entrainer une qualité de dessin inférieure à celle de la recopie, cette stratégie devrait permettre de meilleures performances en termes de rappel et de compréhension, grâce à l'engagement cognitif plus important qu'elle implique. Elle serait donc susceptible d'entrainer un effet de difficulté désirable [22] dont les conséquences serait positives sur l'apprentissage.

Ainsi, nous nous attendons à ce que les étudiant·es de la condition « recopie » produisent des dessins de meilleure qualité que ceux de la condition « mémorisation » (Hypothèse 1). En revanche, nous faisons l'hypothèse que les étudiant·es de la condition

« mémorisation » obtiendront des scores plus élevés en mémorisation et en compréhension des concepts anatomiques, grâce à l'effort de récupération associée à la stratégie « de mémoire » (Hypothèse 2).

Par ailleurs, les capacités visuospatiales jouent un rôle clé dans l'apprentissage de l'anatomie. Des travaux ont montré qu'elles sont prédictives de l'acquisition des connaissances anatomiques. Une revue systématique a mis en évidence une relation significative entre les aptitudes spatiales et l'acquisition des connaissances en anatomie [23]. Une méta-analyse plus récente a également révélé une corrélation positive entre ces deux mesures [24]. Le lien est observé lorsque les performances au test d'anatomie sont mesurées à l'aide de tests ayant une composante visuelle forte, comme la complétion de schémas par exemple. Les capacités visuospatiales pourraient donc avoir une influence sur l'efficacité de la stratégie de dessin. Or, dans les travaux qui concernent cette stratégie, les capacités visuospatiales sont généralement contrôlées mais leur effet n'est que très peu examiné et semble varier selon le contenu d'apprentissage et le contexte. Par exemple, des chercheures [25] ont montré que pour l'apprentissage d'un système mécanique (pompe à vélo), la création de dessin a été plus bénéfique aux étudiant·es ayant de faibles capacités visuospatiales, alors pour l'apprentissage de liaisons chimiques, ce sont les étudiant·es ayant de hautes capacités visuospatiales qui ont bénéficié de la stratégie de dessin. Par ailleurs, une autre étude [26] a montré que les capacités visuospatiales étaient prédictives des performances de transfert mais pas du rappel. Enfin, une étude plus récente [27] a nuancé le lien entre capacités spatiales et performances d'apprentissage, en mettant en avant le fait que ce lien dépend du type de test utilisé pour évaluer l'acquisition des connaissances anatomiques.

Dans cette étude, nous examinons donc l'effet des capacités visuospatiales sur l'apprentissage par le dessin. Nous postulons que les capacités visuospatiales influencent l'efficacité de la stratégie de dessin. Plus précisément, nous suggérons que les étudiant·es ayant des capacités visuospatiales élevées produiront des dessins de meilleure qualité que ceux ayant de faibles capacités visuospatiales (Hypothèse 3). De plus, ces étudiant·es obtiendront des performances supérieures en rappel et en compréhension des connaissances anatomiques apprises (Hypothèse 4).

2 Méthode

2.1 Participants

L'étude a été menée auprès de 84 étudiant·es en deuxième année de kinésithérapie. L'âge moyen était de 22 ans (ET = 3.45). L'échantillon était composé de 49 femmes et 35 hommes. Les étudiant·es ont été répartis en deux groupes : 41 étaient dans la condition « recopie » et 43 dans le groupe « mémorisation ».

Un consentement éclairé a été signé par l'ensemble des participant·es. Le protocole d'étude a été validé par le comité d'éthique de l'Université, et les données ont été anonymisées.

2.2 Matériel

Le contenu pédagogique était présenté sur tablette Dell Latitude 5285 (12.3'') avec stylet, était dynamique et segmenté en étapes (voir Figure 1). L'exercice était composé de deux phases. Une phase de consultation du cours (sans activité de dessin), puis une seconde phase dans laquelle le cours était rejoué. Dans cette seconde phase, il était demandé aux étudiant·es de dessiner le modèle étape par étape. Le matériel consistait en une leçon sur les ligaments longitudinaux du rachis. Celui-ci a été conçu par une formatrice en anatomie spécialisée à la technique du croquis anatomique.

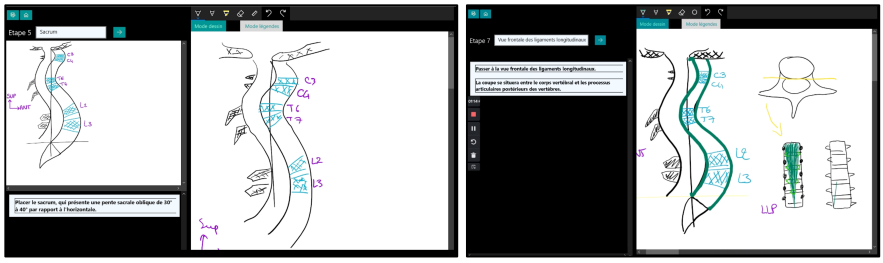


Fig. 1. Copies d'écran des conditions "recopie" (à gauche) et « mémorisation » (à droite).

2.3 Logiciel SKETCH

Chaque participant·e avait à disposition une tablette avec stylet. Une application développée pour le projet contexte à l'étude par l'équipe SHADOC de l'IRISA de Rennes, a été utilisée pour présenter le contenu d'apprentissage ainsi que permettre la réalisation du dessin d'apprentissage. La partie gauche de l'écran était consacrée au cours d'anatomie construit par l'enseignante. Ce cours était segmenté en étapes et les apprenant·es avançaient à leur rythme étape par étape. La partie droite de l'écran était une zone active de dessin dans laquelle les étudiant·es pouvaient réaliser l'activité en fonction de la condition dans laquelle iels étaient affilié·es (voir Figure 1). Une barre d'outils située en haut de cette zone contenait plusieurs fonctionnalités : différents types de crayons avec la possibilité d'ajuster l'épaisseur et choisir la couleur, une gomme, une règle, la possibilité d'annuler ou rétablir une action, et la possibilité de zoomer.

Les apprenant·es ne pouvaient pas revenir en arrière lorsqu'iels avaient passé une étape. De plus, iels ne pouvaient pas modifier des tracés qui avaient été réalisé dans une étape précédente.

2.4 Mesures

Des informations sur les participant·es, leur intérêt pour le dessin et l'anatomie, leurs capacités visuospatiales et leurs connaissances préalables ont été mesurées dans le prétest, en tant que variables contrôles. Les performances d'apprentissage et l'utilité perçue ont été évaluées dans le post-test en tant que variables dépendantes.

Test de connaissances préalables : ce test comportait 6 questions à choix unique. Chaque question était notée sur 1 point pour un total de 6 points.

Capacités visuospatiales : le « Paper Folding Test » de Ekstrom et al. (1976) se compose de deux parties chacune comportant 10 questions à choix multiples. Les participant·es doivent manipuler mentalement le pliage et le dépliage d'une feuille de papier qui a été trouée. Iels choisissent la bonne réponse parmi cinq réponses possibles. Un point est attribué par bonne réponse, un point est soustrait en cas d'erreur ou d'absence de réponse.

Les performances d'apprentissage ont été évaluées à l'aide de plusieurs types de tests : un test de rappel visuel, un test de mémorisation et un test de compréhension. Le test de rappel visuel était une tâche de complétion de croquis. Le croquis du rachis était présenté aux étudiant·es avec des éléments manquants et il leur était demandé de compléter celui-ci. Ce test était évalué avec un score de présence (sur 9 points) et un score de précision (sur 9 points). Le score de présence mesure le nombre d'éléments dessiné par les étudiant·es (1 point est accordé par élément). Le score de précision correspond au nombre d'éléments correctement dessinés, avec 1 point attribué pour chaque élément réalisé avec précision. Le test de mémorisation, composé de 6 items était identique au test de connaissances préalables administré en pré-test. Le test de compréhension, composé de 5 questions à choix unique, évalue la capacité des étudiant·es à appliquer les connaissances acquises dans des situations concrètes, vérifiant ainsi leur compréhension fonctionnelle nécessaire pour la pratique de kinésithérapeute. Ce test était évalué sur 5 points.

L'utilité perçue de la stratégie de dessin pour l'apprentissage : cette variable a été mesurée lors du pré-test ($\alpha = .924$) et du post-test ($\alpha = .942$), à l'aide de 4 items (e.g., « *Je trouve que l'activité de dessin est utile pour l'apprentissage de l'anatomie »)*. Les apprenant·es ont été invité·es à évaluer leur degré d'accord avec chaque item sur une échelle allant de 1 (« pas du tout d'accord ») à 9 (« tout à fait d'accord »).

2.5 Procédure

Au début de l'expérimentation, tous les participant·es ont reçu des informations sur le contexte de l'étude et ont rempli un formulaire de consentement éclairé. L'étude s'est déroulée dans un amphithéâtre d'un institut de formation en kinésithérapie et les étudiant·es étaient accueilli·es par groupe, en fonction de la condition testée. Chaque participant·e était assi·es individuellement à un bureau équipé d'une tablette avec un stylet. Les participant·es ont complété un premier questionnaire visant à évaluer leur intérêt pour l'anatomie et le dessin, leurs compétences visuospatiales, ainsi que leurs connaissances préalables sur les ligaments longitudinaux du rachis. Ensuite, une phase de familiarisation avec le logiciel leur a permis de prendre en main l'outil d'apprentissage.

La phase d'apprentissage a suivi, durant laquelle tous les étudiant·es ont étudié le cours d'anatomie de manière progressive. Les étudiant·es du groupe « recopie » étaient amené·es à réaliser leur dessin d'apprentissage à partir d'un modèle visuel fourni, tandis que le groupe « mémorisation » réalisaient leur dessin de mémoire à partir d'une aide textuelle. L'ensemble des étudiant·es étaient informé·es que leurs connaissances seraient interrogées à l'issue de cette activité.

Après la phase d'apprentissage, un questionnaire post-test a été distribué pour mesurer les performances d'apprentissage et les perceptions des étudiant·es.

3 Résultats

Les analyses ont été effectuées à l'aide du logiciel Jamovi. Le seuil de significativité a été fixé à a = .05.

3.1 Variables contrôles

Avant de tester nos hypothèses, nous avons examiné si les groupes expérimentaux différaient en termes d'âge, de genre, de latéralité, ainsi que d'intérêt et de niveaux perçus en anatomie et en dessin. Les analyses de Khi carré n'ont pas révélé de différence significative entre les conditions en termes de genre $X^2(1, 84) = .720$, $p = .396$, et de latéralité, $X^2(1, 84) = 2.01$, $p = .157$. Les résultats concernant les variables contrôlées d'intérêt et de niveaux perçus sont indiqués dans le tableau 1.

Tableau 1. Statistiques descriptives des variables contrôlées d'intérêt et niveaux perçus.

Mesures	Conditions	M	ET	*t*	ddl	*p*
Intérêt pour l'anatomie	Recopie	7.63	1.04	-.518	82	.606
	Mémorisation	7.74	.902			
Intérêt pour le dessin	Recopie	4.49	2.10	-1.265	82	.209
	Mémorisation	5.09	2.28			
Niveau perçu en anatomie	Recopie	6.66	.728	.421	82	.675
	Mémorisation	6.58	.932			
Niveau perçu sur l'œil	Recopie	5.90	.944	-.114	82	.910
	Mémorisation	5.93	1.26			
Niveau perçu en dessin	Recopie	3.80	1.69	-1.200	82	.233
	Mémorisation	4.30	2.08			

Concernant les capacités visuospatiales, l'analyse du t de Welch ne révèle pas de différence significative entre la condition « recopie » ($M = 7.66$, $ET = 7.67$) et la condition « mémorisation » ($M = 10.2$, $ET = 5.07$), $t(68.9) = -1.81$, $p = .075$

Concernant les connaissances préalables, le t test de Student ne révèle pas de différence significative entre la condition « recopie » ($M = 2.12$, $ET = 1.10$) et la condition « mémorisation » ($M = 2.19$, $ET = 1.10$), $t(82) = -.267$, $p = .790$. Par ailleurs, nous nous sommes assurés que tous les étudiant·es aient peu de connaissances préalables sur le rachis. Deux participant·es ayant eu des scores trop élevés au test de connaissances préalables (> 5 points) ont été considérés comme expert·es et retiré·es des analyses.

Les analyses montrent donc que les groupes sont homogènes pour l'ensemble des mesures contrôles de cette étude. Les distributions sont normales et le jeu de données ne présente pas de valeurs extrêmes.

3.2 Qualité des dessins produits pendant l'apprentissage

Des analyses de régressions linéaires ont été réalisées pour analyser l'effet de la condition d'apprentissage et des capacités visuospatiales sur la qualité des dessins réalisés pendant l'apprentissage.

Concernant le score de présence, le modèle était significatif, $F(3, 80) = 6.16$, $p < .001$, avec un coefficient de détermination $R^2 = .188$. Les résultats ne montrent pas d'effet de la condition sur cette mesure ($\beta = -.130$, $t = -.189$, $p = .850$). L'hypothèse 1 n'est donc pas validée. En revanche, l'analyse révèle un effet des capacités visuospatiales, ($\beta = .076$, $t = 2.184$, $p = .032$). Les étudiant·es ayant de hautes capacités visuospatiales ($M = 31.7$, $ET = .265$) dessinent plus d'éléments que les étudiant·es ayant des faibles capacités visuospatiales ($M = 31.6$, $ET = .285$). Ce résultat est toutefois à nuancer, du fait d'un effet plafond sur les données. L'analyse révèle un effet d'interaction, $\beta = .132$, $t = .062$, $p = .035$. L'effet des capacités visuospatiales semble dépendre de la condition d'apprentissage.

Concernant le score de précision, le modèle était significatif, $F(3, 80) = 15$, $p < .001$, avec un coefficient de détermination $R^2 = .361$. Les analyses montrent un effet de la condition sur le score de précision, ($\beta = -7.063$, $t = -3.598$, $p < .001$), ainsi qu'un effet des capacités visuospatiales, ($\beta = .212$, $t = 2.153$, $p = .034$). En revanche, les résultats ne montrent pas d'effet d'interaction, ($\beta = .008$, $t = .046$, $p = .963$). Les étudiant·es en condition « recopie » ($M = 29.1$, $ET = .756$) ont donc produit des dessins significativement plus précis que les étudiant·es de la condition « mémorisation » ($M = 22.1$, $ET = .750$), conformément à l'hypothèse 1. Par ailleurs, indépendamment de la condition, les capacités visuospatiales semblent favoriser la qualité des dessins produit pendant l'apprentissage. Conformément à l'hypothèse 3, les dessins produits par les étudiant·es ayant de hautes capacités visuospatiales ($M = 27.1$, $ET = .754$) sont plus précis que ceux produits par les étudiant·es ayant de faibles capacités visuospatiales ($M = 24.2$, $ET = .812$).

3.3 Performances d'apprentissage

Des analyses de régressions linéaires ont été réalisées pour analyser l'effet de la condition d'apprentissage et des capacités visuospatiales sur les performances d'apprentissage des étudiant·es.

Concernant le test de mémorisation, les résultats montrent que le modèle est significatif, $F(3, 80) = 2.65$, $p = .054$, et explique 30.1% de la variance des performances ($R = .301$, $R^2 = .095$). Les coefficients des prédicteurs révèlent que la condition d'apprentissage est un facteur significatif ($\beta = .967$, $t = 1.94$, $p = .056$). Conformément à l'hypothèse 2, les étudiant·es de la condition « mémorisation » ($M = 5.93$, $ET = 1.16$) obtiennent des résultats plus élevés que ceux de la condition « recopie » ($M = 5.41$, $ET = 1.34$). De plus, les capacités visuospatiales ont un effet significatif sur la mémorisation ($\beta = .053$, $t = 2.12$, $p = .037$). Les étudiant·es ayant de hautes capacités visuospatiales obtiennent des scores significativement plus élevés (Hypothèse 4). Une analyse

supplémentaire a été réalisée pour tester une interaction entre la condition d'apprentissage et les capacités visuospatiales. Cette interaction n'était pas significative (p = .182).

Concernant la compréhension, le modèle n'était pas significatif, $F(3, 80) = 2.30$, p = .068, avec un coefficient de détermination R^2 = .084. Les résultats montrent que la condition d'apprentissage n'a pas d'effet sur la compréhension (β = -.069, t = -.206, p = .837). L'hypothèse 2 n'est donc pas validée pour cette seconde mesure. En revanche, les capacités visuospatiales ont un effet significatif sur la compréhension, (β = .040, t = 2.385, p = .019). Les étudiant·es ayant des capacités visuospatiales élevées obtiennent des scores plus élevés au test de compréhension (Hypothèse 4). L'interaction entre la condition d'apprentissage et les capacités visuospatiales n'était pas significative (p = .781).

Pour le score de présence relatif à la complétion du croquis (tâche de rappel visuel), le modèle n'était pas significatif, $F(3, 80) = 1.75$ avec R^2 = .059. Ni la condition d'apprentissage (β = -.169, t = -.645, p = .521) ni les capacités visuospatiales (β = .021, t = 1.602, p = .113) n'ont montré d'effet significatif sur le score de présence. Il en est de même concernant le score de précision. Le modèle n'était pas significatif, $F(3, 80) = 1.24$ avec R^2 = .044. Ni la condition d'apprentissage (β = -.409, t = -.687, p = .494), ni les capacités visuospatiales (β = .042, t = 1.396, p = .166) n'a d'effet sur le score de précision. Les deux modèles ne présentent pas d'interaction significative.

Tableau 2. Statistiques descriptives des performances d'apprentissage par condition

Mesures	Condition	*N*	*M*	*ET*
Test de mémorisation	**Recopie**	41	5.41	1.34
	Mémorisation	43	5.93	1.16
Test de compréhension	**Recopie**	41	2.61	.862
	Mémorisation	43	2.56	.825
Score présence	**Recopie**	41	8.83	.543
	Mémorisation	43	8.63	.725
Score de précision	**Recopie**	41	7.85	1.53
	Mémorisation	43	7.58	1.38

3.4 Utilité perçue de la stratégie de dessin pour l'apprentissage de l'anatomie

Une analyse de covariance à mesures répétées (ANCOVA) a été réalisée pour examiner l'effet de la condition d'apprentissage sur les perceptions des étudiant·es concernant la stratégie de dessin. Les résultats n'ont pas montré d'effet principal de la condition, $F(1, 81) = 1.19$, p = .280, ni du moment de la mesure $F(1, 81) = .069$, p = .792. De plus, l'interaction entre la condition d'apprentissage et le moment du test n'était pas significative, $F(1, 81) = .015$, p = .902. De manière descriptive, nous observons des scores particulièrement élevés d'utilité perçue de la stratégie de dessin dès le pré-test. Ces

estimations restent élevées après la mise en situation d'apprentissage (Tableau 3). Globalement, les étudiant·es estiment que le fait de réaliser un dessin est particulièrement utile à l'apprentissage de l'anatomie, et l'activité proposée lors de cette étude n'a pas modifié leur perception.

Tableau 3. Score d'utilité perçue des différentes stratégies de dessin d'apprentissage estimée en pré-test et post-test (Likert de 1 à 9)

Utilité perçue	Conditions	N	M	ET
Pré-test	Recopie	40	6.71	1.91
	Mémorisation	43	7.11	1.42
Post-test	Recopie	41	6.79	1.91
	Mémorisation	43	7.13	1.59

4 Discussion

L'objectif principal de cette étude était de tester deux stratégies d'apprentissage par le dessin. Notre première hypothèse supposait un effet de la recopie sur la qualité des dessins produit pendant la phase d'apprentissage. En accord avec cette hypothèse, les résultats montrent que la qualité des dessins est supérieure en condition « recopie » par rapport à la condition « mémorisation », confirmant notre hypothèse. Ce constat est cohérent avec l'idée que la recopie fourni un modèle visuel direct pendant l'élaboration du dessin, facilitant une reproduction précise des éléments anatomiques. Les étudiant·es qui ont recopié le modèle ont produit des dessins plus complets et plus précis que celleux qui ont généré de mémoire. Cependant, la qualité supérieure des dessins produits en recopie n'a pas conduit à de meilleures performances d'apprentissage. Comme nous le prédisions avec la deuxième hypothèse, c'est la stratégie de dessin réalisé de mémoire qui a eu un effet sur l'apprentissage. En effet, notre deuxième hypothèse est partiellement validée. Les résultats ont montré que l'effort de récupération induite par la condition « mémorisation » a conduit à un meilleur rappel des éléments à apprendre. Ces premiers résultats contredisent l'effet pronostique rapporté dans la littérature, selon lequel la qualité des dessins réalisés pendant l'apprentissage serait un prédicteur des performances d'apprentissage [28]. Dans le contexte de notre étude, la qualité des dessins joue un rôle moins déterminant que l'effort cognitif engagé pendant la tâche d'apprentissage. En revanche, le type de tâche de dessin n'a pas eu d'effet sur la complétion de dessin et la compréhension. Ce résultat pourrait être expliqué par la nature du cours qui correspond principalement à de l'anatomie descriptive. S'agissant de connaissances plus factuelles que fonctionnelles, la nature du cours peut ne pas avoir facilité le transfert direct des connaissances vers la pratique.

Le deuxième objectif de cette étude visait à explorer l'effet des capacités visuospatiales sur la stratégie de dessin et les performances d'apprentissage. Ces aptitudes se sont révélées avoir un effet sur ces mesures, validant nos hypothèses 3 et 4. Les étudiant·es ayant des capacités visuospatiales élevées ont produit des dessins de meilleure qualité, indépendamment du type de tâche de dessin, et ont également obtenu de meilleurs scores de rappel et de compréhension. Ces résultats confirment l'impact des capacités visuospatiales dans les activités d'apprentissage impliquant la manipulation de contenus visuels complexes, tel que la génération de dessins anatomiques.

Les résultats de cette étude suggèrent donc que, dans un contexte d'apprentissage de l'anatomie qui nécessite une représentation visuelle interne, l'activité de dessin en tant que stratégie générative doit être adaptée pour être efficace sur l'apprentissage. L'effort de récupération mobilisé par le fait de dessiner « de mémoire » semble apporter un bénéfice pour les étudiant·es comparativement à une activité de recopie. Par ailleurs, les capacités visuospatiales semblent être un facteur important à prendre en compte dans l'analyse de la stratégie de dessin. Celles-ci étant prédictives des performances d'apprentissage.

Enfin, les résultats relatifs à l'utilité perçue de l'activité de dessin pour l'apprentissage de l'anatomie n'ont pas révélé de différence significative entre les deux stratégies utilisées. Cependant, les scores élevés obtenus tant au pré-test qu'au post-test indiquent que, déjà avant l'étude, l'activité de dessin était considérée comme particulièrement utile pour l'apprentissage de l'anatomie. Cette perception positive est restée stable après la phase d'apprentissage. L'expérience de génération de dessin n'a pas modifié leur conviction quant à l'utilité de cette stratégie pour l'apprentissage des structures du corps humain.

Cette étude présente plusieurs limites. Tout d'abord, l'absence d'un groupe contrôle, sans activité de dessin, ne permet pas de vérifier l'effet de la génération sur les performances d'apprentissage mais cet effet positif a déjà été démontré dans d'autres études du projet Par ailleurs, la condition « mémorisation » repose sur la théorie de la difficulté désirable, qui suggère un effet à long terme. Cependant, notre étude évalue principalement l'apprentissage à court terme. Une mesure différée, à une à deux semaines après l'apprentissage serait nécessaire. De futures recherches sont également nécessaires pour mieux comprendre comment d'autres variables que les capacités visuospatiales peuvent moduler les effets des stratégies de dessin sur l'apprentissage de l'anatomie. Par exemple, d'autres études du projet explorent déjà l'influence de facteurs tels que la charge cognitive, l'intérêt situationnel ou la métacognition. Ces études s'inscrivent dans l'objectif du projet de concevoir un STI, sur tablette, avec une IA capable d'analyser les dessins des étudiants et de fournir un feedback en temps réel lors de leur apprentissage.

Conflit d'intérêts. Les auteur·es déclarent n'avoir aucun conflit d'intérêt.

Remerciements. Ce travail est financé par le projet ANR numéro ANR-21-CE38-0009-01. Nous tenons à remercier nos collaborateurs, Eric Anquetil, Nathalie Girard,

Omar Krichen et Islam Barchouch (IRISA), ainsi que les formateur·rices de l'Institut de formation en kinésithérapie.

Références

1. Chi MTH, Wylie R. The ICAP Framework: Linking Cognitive Engagement to Active Learning Outcomes. Educ Psychol. 2 oct 2014;49(4):219-43.
2. Mayer RE, Fiorella L. Learning as a Generative Activity.
3. Cromley JG, Du Y, Dane AP. Drawing-to-Learn: Does Meta-Analysis Show Differences Between Technology-Based Drawing and Paper-and-Pencil Drawing? J Sci Educ Technol. avr 2020;29(2):216-29.
4. Fiorella L, Zhang Q. Drawing Boundary Conditions for Learning by Drawing. Educ Psychol Rev. sept 2018;30(3):1115-37.
5. Van Meter P, Firetto C. Cognitive model of drawing construction. Learn Vis Disp. 2013;247-80.
6. Van Meter P, Garner J. The Promise and Practice of Learner-Generated Drawing: Literature Review and Synthesis. Educ Psychol Rev. déc 2005;17(4):285-325.
7. Hellenbrand J, Mayer RE, Opfermann M, Schmeck A, Leutner D. How generative drawing affects the learning process: An eye-tracking analysis. Appl Cogn Psychol. nov 2019;33(6):1147-64.
8. Leutner D, Leopold C, Sumfleth E. Cognitive load and science text comprehension: Effects of drawing and mentally imagining text content. Comput Hum Behav. mars 2009;25(2):284-9.
9. Zhang Q, Fiorella L. Learning by drawing: When is it worth the time and effort? Contemp Educ Psychol. juill 2021;66:101990.
10. Haßler B, Major L, Hennessy S. Tablet use in schools: a critical review of the evidence for learning outcomes. J Comput Assist Learn. avr 2016;32(2):139-56.
11. Lai C -H., Yang J -C., Chen F -C., Ho C -W., Chan T -W. Affordances of mobile technologies for experiential learning: the interplay of technology and pedagogical practices. J Comput Assist Learn. août 2007;23(4):326-37.
12. Ciampa K. Learning in a mobile age: an investigation of student motivation. J Comput Assist Learn. févr 2014;30(1):82-96.
13. Furió D, Juan M -C., Seguí I, Vivó R. Mobile learning vs. traditional classroom lessons: a comparative study. J Comput Assist Learn. juin 2015;31(3):189-201.
14. Jamet E, Michinov E. Effects of verbal and visual support on learning by tablet-based drawing. Comput Educ. mai 2022;181:104460.
15. Rochat J, Jamet É, Michinov E. Learning by tablet-based drawing: Effects of verbal support among secondary-school students: L'Année Psychol. 21 juill 2022;Vol. 122(3):513-36.
16. Schmidgall SP, Eitel A, Scheiter K. Why do learners who draw perform well? Investigating the role of visualization, generation and externalization in learner-generated drawing. Learn Instr. avr 2019;60:138-53.
17. Schmidgall SP, Scheiter K, Eitel A. Can we further improve tablet-based drawing to enhance learning? An empirical test of two types of support. Instr Sci. août 2020;48(4):453-74.

18. Borrelli M, Leung B, Morgan M, Saxena S, Hunter A. Should drawing be incorporated into the teaching of anatomy? J Contemp Med Educ. 2018;1.
19. Naug HL, Colson NJ, Donner DG. Promoting metacognition in first year anatomy laboratories using plasticine modeling and drawing activities: A pilot study of the "Blank Page" technique. Anat Sci Educ. juill 2011;4(4):231-4.
20. Naug HL, Colson NJ, Donner D. Experiential Learning, Spatial Visualization and Metacognition: An Exercise with the "Blank Page" Technique for Learning Anatomy. Health Prof Educ. juin 2016;2(1):51-7.
21. Peart DJ. Hand drawing as a tool to facilitate understanding in undergraduate human biology: a critical review of the literature and future perspectives. Stud Sci Educ. 2 janv 2022;58(1):81-93.
22. Bjork EL, Bjork RA. Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. Psychol Real World Essays Illus Fundam Contrib Soc. 2011;2(59-68).
23. Langlois J, Bellemare C, Toulouse J, Wells GA. Spatial abilities and anatomy knowledge assessment: A systematic review. Anat Sci Educ. juin 2017;10(3):235-41.
24. Roach VA, Mi M, Mussell J, Van Nuland SE, Lufler RS, DeVeau KM, et al. Correlating Spatial Ability With Anatomy Assessment Performance: A Meta-Analysis. Anat Sci Educ. mai 2021;14(3):317-29.
25. Bobek E, Tversky B. Creating visual explanations improves learning. Cogn Res Princ Implic. déc 2016;1(1):27.
26. Zhang Q, Fiorella L. Role of generated and provided visuals in supporting learning from scientific text. Contemp Educ Psychol. oct 2019;59:101808.
27. Ritchie HE, Oakes D, Drury H, Ollerenshaw S, Hegedus E. Can drawing instruction help students with low visuospatial ability in learning anatomy? Anat Sci Educ. mars 2023;16(2):252-65.
28. Schwamborn A, Mayer RE, Thillmann H, Leopold C, Leutner D. Drawing as a generative activity and drawing as a prognostic activity. J Educ Psychol. nov 2010;102(4):872-9.

Quand utilité effective et perception de l'utilité diffèrent : le cas d'un jeu sérieux pour l'entraînement post-appareillage d'enfants malentendants

Maëlys Le Magadou[1/2], Bétrancourt Mireille[1], Perez-Fornos Angélica[2], Glück Florent[3], Deriaz Marielle[2], Valentini Gregory[2], Baussand Lou[2], Fritz Matilde[2]

[1] TECFA, Université de Genève, Genève, Suisse
prenom.nom@unige.ch
[2] CURIC, Hôpitaux Universitaires de Genève, Genève, Suisse
prenom.nom@hug.ch
[3] HEPIA, Genève, Suisse
prenom.nom@hepia.ch

Résumé. Le développement du langage oral constitue un réel défi pour le jeune enfant malentendant appareillé, malgré les bénéfices des aides auditives et du suivi logopédique/orthophonique. L'enfant doit quotidiennement être exposé à des stimulations auditives adéquates et encouragé à l'initiative vocale, nécessitant une forte implication familiale. Pour soutenir les familles, le jeu sérieux *FunSpeech* a été conçu selon une conception centrée humain. Il inclut six activités basées sur un principe d'interaction vocale : l'enfant joue au jeu avec sa voix. Cet article présente le protocole individuel quasi-expérimental évaluant tant l'acceptabilité du jeu auprès des familles que son potentiel pour le développement du langage chez l'enfant appareillé. Après des sessions d'entrainement avec *FunSpeech*, une amélioration qualitative des productions vocales et une meilleure maîtrise de la voix sont observées auprès de plusieurs enfants. Cependant, l'acceptabilité parentale du jeu est compromise par une faible utilité perçue. Des pistes pour contourner cet obstacle sont évoquées.

Mots-clés : acceptabilité, langage, surdité, jeu sérieux, réhabilitation.

Abstract. The development of oral language is a real challenge for young hearing-impaired children, despite the proven benefits of hearing aids and speech therapy. Newly implanted children need to be frequently exposed to appropriate auditory stimulation and encouraged to produce controlled sounds, which requires strong involvement of their family environment. *FunSpeech* is a serious game with a human-centered design. It includes six activities based on the principle of vocal interaction: the child plays with his/her own voice. This article presents the study with single-case design protocol used to assess both the game's acceptability to families and its potential for language development in hearing-impaired children. After the training sessions with *FunSpeech*, we observed a qualitative improvement in vocal productions and better voice control on specific parameters in several participants. However, the game's acceptability is compromised by the poor perception of parents regarding of its usefulness. Recommendations for circumventing this obstacle are provided.

Keywords: acceptance, language, hearing loss, serious games, rehabilitation

1 Introduction

Le langage oral se développe de manière spontanée chez le bébé entendant, tant par la mobilisation de ses capacités que par les sollicitations de son environnement. Chez le jeune enfant malentendant, ce développement est entravé par le déficit auditif [1]. Dans plus de 85% des cas, l'enfant malentendant évolue au sein d'une famille entendante, où rapidement une discordance entre ses capacités et son environnement linguistique apparait : l'enfant ne perçoit pas le son ou le perçoit de manière déformée, l'empêchant de le reproduire correctement. L'ajustement de ses productions vocales est également gêné par un feedback auditif insuffisant. L'acquisition des compétences prélinguistiques et phonologiques est ainsi fragilisée, tout comme les échanges de l'enfant avec ses partenaires de communication. De plus, une surdité survenue dans la petite enfance peut affecter le développement cérébral, le potentiel cognitif ou encore les compétences sociales de l'enfant.

Dans une perspective de réhabilitation, les bénéfices des aides auditives (i.e., prothèse auditive, implant cochléaire, implant à ancrage osseux) font aujourd'hui consensus. Bien que fonctionnelles, les capacités auditives restituées demeurent limitées et la mise en place d'un accompagnement logopédique/orthophonique conséquent est indispensable. Afin d'expliciter le langage, les logopédistes[1] proposent une palette d'activités, de l'éveil sonore à la construction de phrases, articulée autour d'une scénarisation ludique soutenant l'engagement de l'enfant. Parallèlement, ce suivi se construit en étroite collaboration avec les parents, dans une perspective de préservation de leur rôle crucial pour le développement du langage de leur enfant. Bien que déterminante, l'implication parentale peut cependant être fragilisée par le choc du diagnostic de surdité et les parents peuvent rencontrer des difficultés à poursuivre à domicile et de manière adéquate l'entraînement réalisé en logopédie.

Afin d'accompagner les familles, un jeu sérieux a été conçu pour soutenir le développement de la production du langage chez l'enfant malentendant appareillé. *FunSpeech* s'inscrit dans la continuité des travaux de Bozelle [2], ayant souligné tant la pertinence de l'utilisation d'un jeu sérieux pour améliorer la compréhension du langage chez l'enfant appareillé que la nécessité de prendre en compte l'acceptabilité parentale des solutions numériques destinées à l'entrainement post-appareillage.

Cet article présente brièvement les approches de l'acceptabilité mobilisées pour l'étude, un aperçu de solutions numériques soutenant la production du langage et le jeu *FunSpeech*. Les premiers résultats d'une étude avec protocole quasi-expérimental seront détaillés, puis l'article se conclura sur une discussion et des perspectives d'amélioration du jeu.

[1] La profession de logopédiste en Suisse correspond à celle des orthophonistes en France.

2 État de l'art

Les approches de l'acceptabilité s'intéressent essentiellement aux représentations des personnes vis-à-vis d'une technologie *a priori*, dans une perspective de prédiction d'un futur comportement d'usage [3]. Sur la base des recherches portant sur la prédiction du comportement, le *Technology Acceptance Model* (TAM) proposé par Davis [4] postule que le comportement d'usage d'une technologie est déterminé par l'intention comportementale, elle-même déterminée par l'attitude (i.e., favorable ou non) vis-à-vis de la technologie. Cette dernière est influencée par deux variables : (i) d'une part l'*utilité perçue*, c'est-à-dire le degré selon lequel une personne pense que l'utilisation d'une technologie peut améliorer sa performance et (ii) d'autre part la *facilité d'usage perçue*, c'est-à-dire le degré selon lequel une personne pense que l'utilisation de la technologie ne nécessitera pas/peu d'efforts. Cette seconde variable aurait également un effet sur l'utilité perçue, de sorte qu'un système considéré comme simple d'utilisation serait également estimé plus utile (Voir **Fig. *1***).

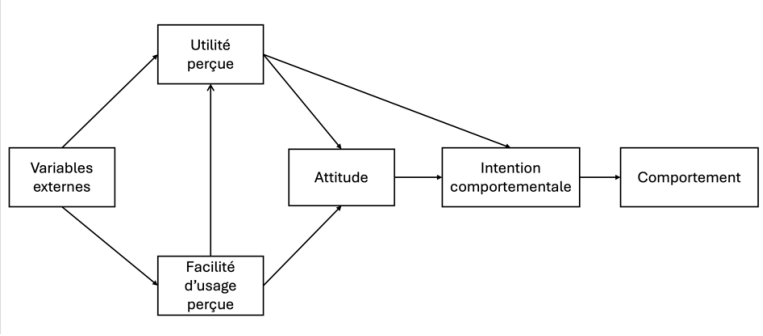


Fig. 1. Modélisation du TAM.

Dans la lignée de ces travaux, Venkatesh et collaborateurs [5] proposent le modèle UTAUT2 (*Unified Theory of Acceptance and Use of Technology*), dans lequel le comportement d'usage est également déterminé par l'intention d'usage, influencée elle-même par plusieurs facteurs :

i. L'*attente de performance*, qui renvoie au degré selon lequel la personne pense que l'utilisation de la technologie va soutenir sa performance ;
ii. L'*attente d'effort*, qui renvoie au degré selon lequel la personne estime que l'utilisation de la technologie nécessitera peu d'efforts ;
iii. L'*influence sociale*, qui est définie comme le degré selon lequel une personne perçoit que les personnes importantes pour elle pensent qu'elle devrait utiliser le système ;
iv. *Les conditions facilitatrices*, qui désignent le degré selon lequel un individu croit qu'une organisation et/ou une infrastructure technique existe comme support à l'utilisation d'une technologie ;
v. La *motivation hédonique*, qui renvoie au plaisir suscité par l'utilisation d'une technologie ;
vi. La *valeur pécuniaire*, qui renvoie au rapport coût/bénéfice du système ;

vii. Les *habitudes*, qui désignent les comportements préalables des individus avec la technologie.

Chaumon [6] propose une approche située de l'acceptabilité, considérant le processus d'adoption d'une technologie selon un continuum, articulé autour de deux perspectives : (i) l'*acceptabilité* d'une part, qui renvoie aux impressions des personnes sur la solution avant usage ou lors des premières utilisations et (ii) l'*acceptation* d'autre part, qui vise l'étude du vécu des personnes après usage ainsi que l'appréciation des écarts entre les attentes initialement formulées et les qualités de la solution retenues. La recherche présentée s'inscrit dans cette approche, puisqu'il est essentiel dans un contexte de réhabilitation auditive de considérer tant les facteurs qui initient l'usage d'une solution numérique en complément du suivi logopédique que ceux qui supportent les efforts d'entrainement jusqu'à l'atteinte d'un objectif de réhabilitation.

Dès les années 1990, des solutions numériques ont été proposées pour initier la production vocale chez l'enfant présentant des besoins spécifiques (i.e., perte auditive, trouble du spectre autistique, trouble de la parole), dans une perspective de développement du langage oral. Bien qu'aujourd'hui inutilisable faute de mise à jour, *SpeechViewer* proposait des activités basées sur un principe d'interaction vocale, délivrant un feedback visuel immédiat après chaque production de l'enfant (e.g., une grenouille qui saute). Pratt et collaborateurs [7] ont montré qu'un entrainement répété avec *SpeechViewer* sur une période de 4 mois avait amélioré la production de phonèmes chez quatre enfants malentendants appareillés d'âge préscolaire, soulignant l'intérêt de la solution comme complément au suivi logopédique. S'inspirant de ce premier programme, Rodríguez et collaborateurs [8] ont conçu la solution *PreLingua* qui entraîne les premières étapes du développement langagier par la maîtrise des paramètres de la voix (i.e., intensité, souffle, temps de phonation, hauteur). Menant une étude auprès de 27 personnes présentant des troubles de la parole, les chercheurs ont montré qu'un entrainement de 12 semaines avec *PreLingua* permettait d'observer des améliorations de la maîtrise de l'intensité de la voix et de la gestion du souffle chez 15 sujets. Plus récemment, Navarro-Newball et collaborateurs [9] ont proposé le jeu sérieux *Talking to Teo*, qui s'adresse spécifiquement aux enfants malentendants porteurs d'implants cochléaires de 3 à 11 ans. Scénarisé autour d'une histoire à découvrir avec le parent, le jeu propose une série d'exercices entrainant la production de phonèmes utiles en espagnol. Les premiers essais du jeu ont permis de recueillir des retours positifs auprès des enfants.

Bien qu'offrant de nouvelles perspectives pour l'entrainement post-appareillage de l'enfant malentendant, ces initiatives restent limitées à un usage accompagné par le thérapeute, à une langue définie ou à l'entrainement de certains paramètres de la voix. De plus, les éléments de conception sont peu argumentés et les méthodes d'évaluation de leur efficacité et de leur acceptabilité auprès des familles demeurent sommaires. Dans ce contexte, le projet *FunSpeech* a été initié.

3 Conception du jeu *FunSpeech*

Disponible sur tablette tactile, *FunSpeech* résulte d'une conception centrée humain impliquant des cliniciens (i.e., logopédistes et ingénieur-e-s[2]), des chercheurs aux compétences et perspectives pluridisciplinaires (i.e., ingénierie clinique, psychologie, ergonomie, informatique) et des familles. Cette démarche itérative prenait en compte tant les enjeux cliniques de la réhabilitation auditive que les compétences des enfants et des familles. Les activités du jeu *FunSpeech* sont conçues selon une approche modulaire du langage, le considérant comme composé de sous-systèmes relativement autonomes. Chaque activité vise ainsi à expliciter un paramètre de la voix à l'enfant (i.e., rythme, intensité et hauteur), par des dynamiques ludiques, des interactions basées sur la production vocale [10] et des feedback visuels immédiats (e.g., disparition d'un animal, apparition d'une récompense). La conception du jeu s'inscrit dans une approche didactique de l'éducation auditive basée sur la répétition d'exercices spécifiques [2] et respecte des principes pédagogiques visant l'engagement de l'enfant (i.e., feedback immédiat, progression visible et mécanique endogène) [11] ainsi que des critères ergonomiques axés sur la facilité d'utilisation et l'expérience positive de jeu [12]. Si les mécaniques de jeu ont été conçues pour de jeunes enfants (i.e., entre 2 à 4 ans), l'usage implique une médiation des parents notamment par l'explicitation des consignes, le guidage vers la compréhension des erreurs ou simplement l'interaction autour des activités.

Concrètement, l'enfant accompagné de son parent choisit son avatar personnel puis accède aux six activités proposées par *FunSpeech* via un menu principal, visible sur la **Fig. 2**.



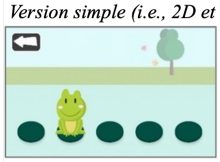
Fig. 2. Menu du jeu *FunSpeech.*

Les activités se distinguent en termes de conception visuelle et de niveau de difficulté. La **Fig. 3** présente les modalités de chaque activité.

2 L'ingénieur-e clinique assure le suivi régulier de l'enfant implanté, tant sur un niveau technique (gestion de l'implant cochléaire) que sur un niveau clinique (réalisation de test standardisé de perception de la parole).

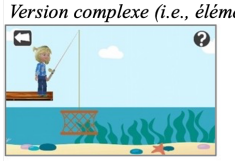
Jeux autour du rythme de la voix

Version simple (i.e., 2D et rythme déterminé par l'enfant) : La grenouille



Le personnage avance lorsque l'enfant fait un son. Il obtient une récompense à la fin du parcours.

Version complexe (i.e., éléments en 3D et rythme imposé à l'enfant) : Les poissons



L'enfant aide le personnage à pêcher en réalisant un son lorsqu'un poisson atteint le panier. Le rythme d'apparition des poissons est modulé par la réussite de l'enfant.

Jeux autour de l'intensité de la voix

Version simple : Le singe



Pour faire disparaître les animaux, l'enfant doit moduler l'intensité de ses productions (i.e., faible, moyenne ou forte). Chaque animal représente une intensité par sa taille et une jauge visuelle guide l'enfant.

Version complexe : Les ours



Un ours se balade en forêt et l'enfant doit l'appeler en modulant l'intensité de sa voix : Par exemple, si l'ours est au premier plan (i.e., « près ») l'enfant doit chuchoter pour le saluer, tandis que si l'ours est en arrière-plan (i.e., « loin ») l'enfant doit crier pour l'appeler.

Jeu autour de la fréquence de voix : *Jeu de l'hélicoptère*



Lorsque l'enfant fait un son aigu, l'objet monte. A l'inverse, la production de sons graves permet de faire descendre l'objet. Le but est de découvrir progressivement une image.

Jeu autour de la répétition de phonèmes : *Jeu du clown*



L'enfant doit écouter le son produit par le personnage (i.e., [a], [i], [u]), puis essayer de le reproduire. Si le son produit correspond au son cible, un animal apparait à l'écran, le clown est heureux et l'enfant-compagnon donne un feedback positif. Dans le cas contraire, le clown est triste et l'enfant-compagnon donne un feedback négatif. Le scénario reprend par un nouveau modèle de l'enfant-compagnon.

Fig. 3. Résumé des jeux et visualisation de l'interface.

Dans le cadre de l'étude clinique, des tablettes tactiles équipées du jeu *FunSpeec*h étaient prêtées aux familles après une première démonstration.

4 Étude clinique

Sur la base des modèles d'acceptabilité évoqués précédemment, les questions de recherche étaient les suivantes :

- **Q1** : Quels sont les effets de l'utilisation du jeu *FunSpeech* sur la progression langagière de l'enfant malentendant appareillé ?
- **Q2** : Si les modèles d'acceptabilité insistent sur l'importance de l'utilité perçue, dans quelle mesure la perception des parents (i.e., *utilité perçue*) et l'acceptabilité de la solution reflètent-elles les performances effectives de l'enfant (i.e., *utilité effective*) ?

4.1 Méthode

Les participants, deux filles et neuf garçons, étaient suivis par le Centre Universitaire Romand d'Implants Cochléaires des Hôpitaux Universitaires de Genève. Les enfants étaient âgés d'un an à 5 ans (Moyenne = 2.6 ans ; Écart-type = 17,6 mois). Tous les participants présentaient un déficit auditif et étaient porteurs d'aide auditive au moment de l'inclusion (durée de port de 1 jour à 36 mois, moyenne 11,4 mois de port). L'étude impliquait également la participation de 14 parents (8 femmes, 6 hommes, âge moyen 39.3 ans, de 20 à 46 ans). Les familles participaient volontairement à l'étude, via une procédure de consentement éclairé. Sept familles disposaient d'une tablette tactile et trois enfants avaient une expérience régulière d'utilisation, avec un accompagnement parental ou en autonomie, pour la réalisation d'activités distrayantes.

4.2 Design de l'étude

L'étude a utilisé un protocole individuel quasi-expérimental, permettant d'appliquer une démarche expérimentale à un nombre limité (voire unique) d'individus, en respectant une variation contrôlée de condition expérimentale et la répétition de mesures au cours du temps [13].

Le protocole a été déployé d'août 2020 à juillet 2021[3]. Chaque enfant constituait sa propre mesure témoin, expérimentant successivement deux conditions expérimentales : une session en condition *contrôle*, où il bénéficiait de son suivi logopédique habituel (i.e., A1 et A2), et une session dite *intervention*, où il disposait du jeu *FunSpeech* à domicile en complément des séances de logopédie (i.e., B1 et B2). Des sessions de jeu en co-usage parent-enfant d'environ 15 minutes par jour étaient recommandées pendant la période de prêt. La durée et la fréquence d'utilisation du jeu étaient déterminées à partir d'un système d'enregistrement automatique. S'inspirant d'un plan ABAB, l'alternance des sessions n'était pas randomisée mais un tirage au sort déterminait la première condition expérimentée par l'enfant. La ligne de base du niveau de langage de l'enfant était évaluée en *t*0, puis après chaque session de 45 jours et selon la même méthode, comme illustré sur la Fig. 4.

[3] Le protocole a été approuvé par la Commission Cantonale d'Éthique de la Recherche sur l'être humain de Genève (ID BASEC 2019-02138).

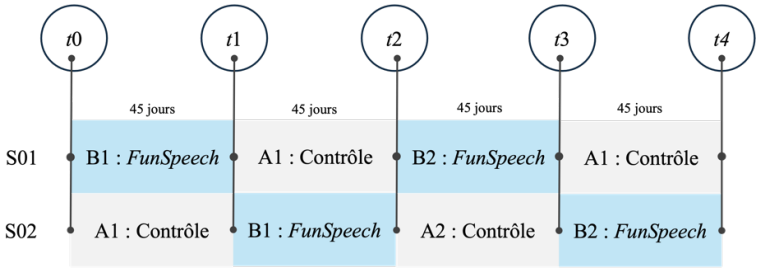


Fig. 4. Évaluations et alternance des conditions

L'organisation conventionnelle du suivi clinique a déterminé la durée des sessions et la fréquence des évaluations, visant tant l'appréhension de tout effet du jeu sur le développement langagier que l'adhésion des familles au protocole. Chaque évaluation était menée lors d'une séance de logopédie filmée d'environ 30 minutes au déroulement standardisé, incluant deux méthodes (voir **Fig. 5**) :

1. **Une séance de jeu libre**, via laquelle les productions spontanées de l'enfant et ses interactions avec son parent étaient relevées à partir d'un matériel standardisé (i.e., ferme, animaux et personnages) [14] ;
2. **Une session avec *FunSpeech***, où l'enfant jouait avec une version alternative conservant les dynamiques de jeu de la version prêtée mais avec de nouveaux graphismes.



Jeu libre



Jeu avec FunSpeech

Fig. 5. Configuration des deux méthodes d'évaluation

Dans cet article, l'effet de *FunSpeech* a été évalué sur la base de deux indicateurs :

(i) L'inventaire phonétique relevé via la méthode « *Jeu libre* », c'est-à-dire le nombre de phonèmes différents maîtrisés par l'enfant sur un répertoire défini de 31 phonèmes ;

(ii) Les réussites aux objectifs ludiques de chaque activité lors des cinq « *Sessions avec FunSpeech* » (i.e., faire avancer l'animal, faire disparaitre un animal, faire voler l'avion).

L'acceptabilité parentale du jeu était évaluée à partir d'un questionnaire de 14 items, distribué aux parents à deux reprises, avant l'utilisation du jeu et après les périodes d'essai de *FunSpeech*. Présentés dans le Tableau 1, les items étaient conçus afin d'évaluer cinq dimensions d'influence de l'acceptabilité selon le modèle UTAUT2 (i.e., précédemment détaillé), l'attitude et l'intention d'usage. La dimension « *Valeur pécuniaire* » était exclue en raison de la gratuité du jeu et le facteur « *Habitudes* » était

décliné sous un format de cases à cocher, de sorte que les parents renseignaient la possession d'une tablette tactile familiale, sa fréquence d'utilisation par l'enfant et les modalités d'utilisation en termes d'accompagnement (e.g., seul ou accompagné) et d'activités réalisées. Ces éléments sont résumés dans la section précédente.

Tableau 1. Résumé des items du questionnaire d'acceptabilité

Dimensions	Items
Attente de performance	Le jeu *FunSpeech* va aider mon enfant à produire des sons. *FunSpeech* permettra à mon enfant de développer son langage.
Attente d'effort	Le jeu me semble facile à utiliser. Ce jeu me semble difficile à utiliser. (i.e., item inversé)
Influence sociale	Je pense qu'il est déconseillé d'utiliser une tablette tactile avec mon enfant. Mon entourage pense qu'il est bien d'utiliser ce jeu pour aider mon enfant.
Conditions facilitatrices	Je sais qui contacter si je rencontre des difficultés avec le jeu. Je pense disposer des compétences nécessaires pour utiliser le jeu avec mon enfant à domicile
Motivation hédonique	Utiliser *FunSpeech* sera une activité amusante pour mon enfant. Ce jeu va motiver mon enfant
Attitude	Je suis favorable à l'utilisation de jeu comme *FunSpeech* pour les enfants. *FunSpeech* est une bonne idée.
Intention d'usage	J'ai l'intention d'utiliser le jeu avec mon enfant après l'étude clinique. Après l'étude clinique, je pense chercher un jeu équivalent à *FunSpeech* pour mon enfant.

Le parent devait déclarer son degré d'accord avec chaque item sur une échelle de Likert en sept points, allant de 1 « *Pas du tout d'accord* » à 7 « *Tout à fait d'accord* », comme présenté ci-dessous.

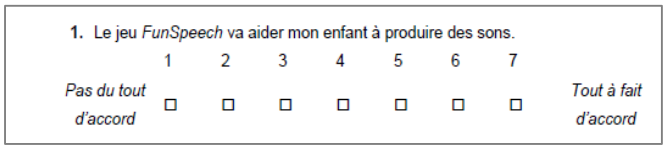


Fig. 6. Exemple d'une affirmation pour l'*Attente de performance*

Un score moyen était déterminé à partir des réponses données aux deux items associés à chaque facteur, à l'attitude et à l'intention d'usage. Ces deux dernières dimensions ne

seront pas détaillées dans cet article. Le questionnaire délivré en post-usage comprenait des encarts complémentaires sous chaque item, permettant de collecter des données qualitatives additionnelles sous la forme de commentaires.

4.3 Résultats

Utilisation du jeu *FunSpeech* lors des sessions d'essai
Les données exposées dans le Tableau 2 permettent d'observer, sur une période de 42 jours, une utilisation faible à modérée du jeu *FunSpeech* lors de la première session d'essai, puis une diminution de celle-ci lors de la seconde session. L'étendue indique également une très grande variabilité de l'utilisation entre les familles.

Tableau 2. Utilisation du jeu en jour et en session selon les sessions B1 et B2

	Nombre de jours d'utilisation		Nombre de sessions réalisées	
	B1	B2	B1	B2
Moyenne	14.5	9.1	18.2	9.7
Étendue	5 - 26	2 - 18	5 - 44	2 - 18

Le temps de jeu moyen est cohérent avec les observations précédentes, comme résumé dans le Tableau 3.

Tableau 3. Temps de jeu moyen et étendue en fonction des sessions B1 et B2 (Format en heure:minute)

	Temps de jeu	
	B1	B2
Moyenne	03:40	02:27
Étendue	01:24 – 09:23	00:41 – 06:07

Q1 : Quels sont les effets de l'utilisation du jeu *FunSpeech* sur la progression langagière de l'enfant malentendant appareillé ?
Afin de faciliter la lecture des données en lien avec l'acquisition du langage, deux *groupes* d'enfants ont été déterminés selon l'ordre des conditions expérimentées au cours du protocole.

Inventaire phonétique durant les séances de jeu libre
Au sein du groupe BABA ayant disposé du jeu *FunSpeech* entre les mesures *t0* et *t1* et les mesures *t2* et *t3*, l'analyse descriptive de l'inventaire phonétique maîtrisé par chaque enfant (Voir **Fig. 7**) permet de relever une extension du répertoire au fil de l'étude pour trois enfants :

- pour S04, plus spécifiquement après les périodes d'essai du jeu *FunSpeech* ;
- pour S09, surtout après la première période d'essai du jeu ;
- pour S13, surtout après la seconde session d'essai du jeu *FunSpeech.*

Il n'y a pas de progression régulière pour les enfants S01 et S02 : d'importantes variations du nombre de phonèmes maîtrisés sont relevées au fil des évaluations.

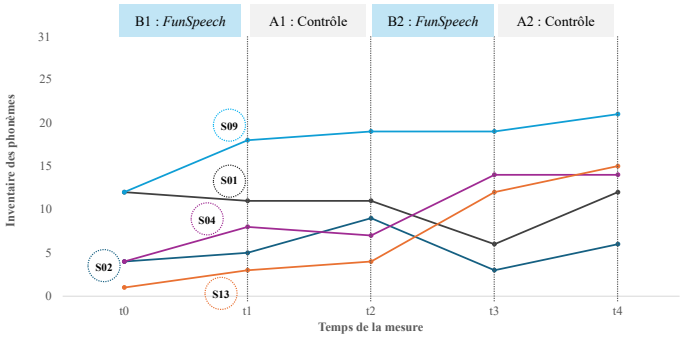


Fig. 7. Représentation visuelle du répertoire de phonèmes maitrisés par l'enfant selon les conditions expérimentales.

Au sein du groupe ABAB ayant disposé du jeu *FunSpeech* entre les mesures *t1* et *t2* et les mesures *t3* et *t4*, une extension des répertoires phonétiques est également observée (Voir **Fig. 8**) pour quatre enfants du début de l'étude à la fin de l'étude :

- Le répertoire des phonèmes de l'enfant S05 est quasi-complet à la fin de l'étude (i.e., 27/31 phonèmes) ;
- Une modeste extension de l'inventaire maîtrisé par l'enfant S07 est relevée, malgré une importante variabilité au fil des évaluations ;
- Chez les enfants S12 et S15, deux importantes évolutions sont visibles après les sessions d'utilisation du jeu *FunSpeech* à domicile ;

Pour les enfants S10 et S16, d'importantes variations sont observées selon les sessions.

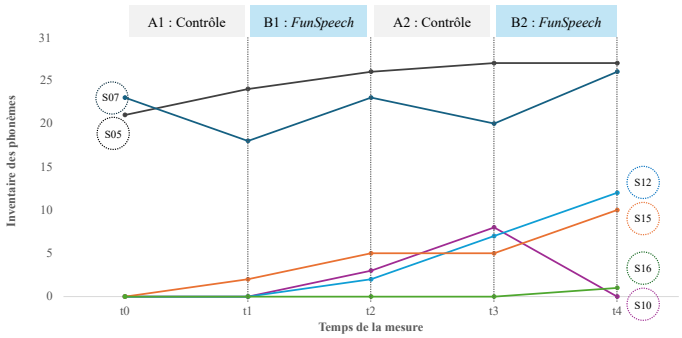


Fig. 8. Représentation visuelle du répertoire de phonèmes maitrisés par l'enfant selon les conditions expérimentales.

En résumé, l'évaluation descriptive des répertoires phonétiques permet de relever une amélioration qualitative du niveau de langage pour sept enfants sur 11 et pour six d'entre eux, ces extensions étaient observées après les deux périodes d'utilisation du jeu *FunSpeech*. Une préservation des améliorations d'une mesure à une autre est également reportée pour la majorité des participants.

Sessions de jeu avec Funspeech

La réussite aux objectifs ludiques du jeu *FunSpeech* relevée au cours des évaluations indique que tous les enfants ont progressé quant à leur maîtrise de leur voix (Voir Tableau 4 et Tableau 5). Ainsi, neuf enfants parvenaient à faire avancer l'animal avec leur voix après l'entrainement avec *FunSpeech*, alors que seulement quatre enfants y parvenaient au début de l'étude. Concernant le jeu portant sur l'intensité de la voix, seul un enfant maîtrisait le principe de jeu au début de l'étude. Après l'entrainement, sept enfants parvenaient à faire disparaitre un animal de manière autonome ou guidée. Enfin, huit enfants jouaient de manière autonome ou guidée avec le jeu portant sur la fréquence de la voix à la fin de l'étude, alors que seulement quatre enfants y parvenaient au début de l'étude.

Tableau 4. Réussites aux objectifs ludiques (cellules colorées en vert) lors des sessions d'évaluation avec *FunSpeech* au cours du temps, groupe BABA ; r. = jeu du rythme ; i. = jeu de l'intensité ; f. = jeu des fréquences ;

				B1			A1			B2			A2		
	t0			*t1*			*t2*			*t3*			*t4*		
	r.	i.	f.	r.	i.	f.	r.	i.	f.	r.	i.	f.	r.	i.	f.
S01		X		X	X			X	X					X	
S02															
S04	X	X	X	X	X										
S09		X													
S13	X	X	-	X	-	-		X	X		-			X	X

Tableau 5. Réussites aux objectifs ludiques (cellules colorées en vert) lors des sessions d'évaluation avec *FunSpeech,* groupe ABAB ; r. = jeu du rythme ; i. = jeu de l'intensité ; f. = jeu des fréquences ;

				A1			B1			A2			B2		
	t0			t1			t2			t3			t4		
	r.	i.	f.	r.	i.	f.	r.	i.	f.	r.	i.	f.	r.	i.	f.
S05	X	X	X		X			X			-	-			
S07		X													
S10	-	-	-	X	X	-					-	-		-	
S12	-	-	-				X	-	X	X		X	X		
S16	-	-	-		X	X			X						X

Pour l'ensemble des enfants, ces résultats indiquent tant une acquisition des principes d'interaction du jeu au fil de l'étude qu'une meilleure maîtrise de leurs productions vocales.

Q2 : Si les modèles d'acceptabilité insistent sur l'importance de l'utilité perçue, dans quelle mesure la perception des parents (i.e., utilité perçue) et l'acceptabilité de la solution reflètent-elles les performances effectives de l'enfant (i.e., utilité effective) ?

Avant usage, les scores moyens relevés pour chaque dimension indiquaient une acceptabilité satisfaisante du jeu auprès des familles (Voir **Fig. 9**) : *a priori* les parents estimaient que le jeu *FunSpeech* pourrait bénéficier à leur enfant au niveau langagier, qu'il serait facile d'utilisation et source de motivation. Ils considéraient également disposer des compétences nécessaires pour utiliser la solution. En revanche, ils exprimaient plus de modération concernant l'utilisation d'un écran avec un jeune enfant.

La comparaison des scores moyens obtenus avant et après usage du jeu pour chaque dimension de l'acceptabilité permet de relever que :

i. Le score moyen pour le facteur *Influence sociale* augmente après l'utilisation du jeu (+ 0.4) ;
ii. Les scores moyens obtenus pour les dimensions *Attente d'effort* (- 0.3) et *Conditions facilitatrices* (- 0.2) baissent légèrement ;
iii. Les scores moyens pour les dimensions *Attente de performance* (- 1.6) et *Motivation* (- 1.7) connaissent une forte baisse.

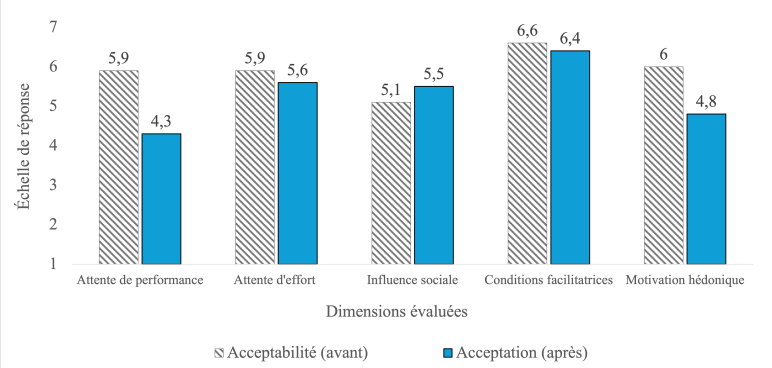


Fig. 9. Scores moyens par dimension de l'acceptabilité avant et après utilisation de *FunSpeech*

Après les essais du jeu *FunSpeech*, la forte diminution des scores moyens obtenus pour les dimensions *Attente de performance* et *Motivation hédonique* indique, selon les représentations des parents, que le jeu n'apporte pas (ou pas suffisamment) de bénéfice pour le développement du langage et qu'il n'est pas (ou pas suffisamment) motivant pour l'enfant. L'analyse des commentaires additionnels confirme cette interprétation :

« Je ne vois pas de lien direct avec la formation de mots. »

« L'amusement était top mais [mon enfant] se lassait rapidement, il changeait de jeu très souvent ».

5 Discussion

Dans une perspective d'intégration du jeu *FunSpeech* au processus de prise en charge post-appareillage, cette recherche s'intéressait à l'évaluation de son potentiel pour le développement langagier du jeune enfant malentendant appareillé et aux facteurs influençant son acceptabilité auprès des parents. Les résultats obtenus révèlent deux informations importantes.

En premier lieu, l'évaluation descriptive des répertoires phonétiques et l'analyse de la réussite aux objectifs ludiques révèlent une progression sensible du niveau langagier pour 7/11 enfants, notamment après les périodes d'entrainement avec le jeu à domicile et ce, malgré une utilisation faible à modérée par les familles. Toutefois, cette progression est particulièrement hétérogène selon les enfants, et s'observe également après des périodes dites « *contrôle* ». Il n'est ainsi pas possible d'affirmer que la tendance à la progression soit liée à l'utilisation du jeu lors des sessions d'essai.

En second lieu, les tendances à la progression n'ont pas été perçues par les parents ou n'ont pas été attribuées aux séances de jeu avec *FunSpeech*. Le recueil de l'acceptation parentale après l'utilisation du jeu à domicile révèle effectivement que les parents déclaraient que le jeu *FunSpeech* ne soutenait pas le développement du langage chez leur enfant (*Attente de performance*) et qu'il n'était pas suffisamment source de motivation pour leur enfant (*Motivation*). En d'autres termes, on observe un écart entre l'utilité effective et l'utilité perçue. De plus, on note une diminution de l'utilisation moyenne de *FunSpeech* entre les deux sessions de prêt, ce qui conforte l'hypothèse d'un lien entre les perceptions parentales et l'utilisation du jeu par les familles.

Dans une perspective de poursuite du suivi logopédique à domicile, il est pourtant essentiel que *FunSpeech* soit perçu comme « *utile* » par les parents puisqu'ils sont à l'initiative des sessions d'entrainement. Les bénéfices potentiels des sessions avec le jeu sur le développement du langage de l'enfant appareillé sont ainsi déterminés par une perception continue de l'utilité du jeu, en cohérence avec les résultats de Bozelle [2]. Une solution possible serait de proposer des outils de suivi des actions de l'enfant au sein du jeu de type tableau de bord, afin que ses parents visualisent son évolution au fil des sessions. Cette amélioration s'accompagne d'une multitude de défis à explorer dans de futures études.

Concernant les limites, le recrutement de parents volontaires pour l'étude clinique pourrait induire un biais dans l'échantillonnage, avec des parents possiblement plus favorables à une solution numérique que l'aurait été un échantillon de parents tout-venant. Néanmoins, plusieurs parents-participants ont mentionné leur crainte concernant l'exposition aux écrans de leur jeune enfant, ce qui suggère qu'il ne s'agissait pas d'une population particulièrement favorable à l'utilisation de *FunSpeech*. D'autre part, il serait risqué de généraliser les résultats obtenus auprès de cet échantillon, d'autant plus dans le cadre de la réhabilitation auditive où les progrès de l'enfant sont influencés

par une pluralité d'autres facteurs. Les résultats soulignent effectivement d'importantes différences inter-individuelles, dont les facteurs d'influence n'ont pas été déterminés.

6 Conclusion et perspectives

Cette recherche souligne le potentiel des protocoles individuels quasi-expérimentaux pour l'évaluation de solutions numériques comme *FunSpeech*. S'intégrant dans un contexte global de prise en charge, comme la réhabilitation auditive d'enfants malentendants, ces protocoles offrent tant un cadre méthodologique rigoureux qu'adaptable aux contraintes des expérimentations dans un milieu clinique. Néanmoins, il s'agit également d'en mentionner les limites, d'autant plus marquées dans cette étude que les participants présentaient des profils hétérogènes : les tendances observées sont ainsi relativement faibles et non-généralisables. Les protocoles ABAB présupposent également des « *ruptures* » de progression entre les périodes « *intervention* » et « *contrôle* », comme lors d'un traitement médicamenteux, qui n'étaient pas recherchées dans cette étude sur le développement langagier de l'enfant. Les prochaines recherches s'orientent d'une part, sur la conception d'outils de suivi pour améliorer la perception « *juste* » de l'utilité effective et d'autre part, sur les possibles effets de ces améliorations sur l'acceptabilité et l'utilisation du jeu *FunSpeech*.

Remerciements

Les auteurs remercient les familles qui ont participé à cette étude.

Références

1. Nicholas, J.G., and Geers, A.E.: Spoken Language Benefits of Extending Cochlear Implant Candidacy Below 12 Months of Age, Otology & Neurotology 34(3), (2013).
2. Bozelle, C.: Evaluation d'un environnement informatisé pour la réhabilitation d'enfants porteurs d'implants cochléaires: utilisation du jeu comme outil d'apprentissage, (2014).
3. Terrade, F., Pasquier, H., Reerinck-Boulanger, J., Guingouain, G., and Somat, A.: L'acceptabilité sociale : la prise en compte des déterminants sociaux dans l'analyse de l'acceptabilité des systèmes technologiques, Le travail humain Vol. 72(4), pp. 383-395 (2009).
4. Davis, F.D.: Technology acceptance model: TAM, Al-Suqri, MN, Al-Aufi, AS: Information Seeking Behavior and Technology Adoption205, pp. 219 (1989).
5. Venkatesh, V., Thong, J.Y., and Xu, X.: Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology, MIS quarterly, pp. 157-178 (2012).
6. Chaumon, M.-E.B.: L'acceptation située des technologies dans et par l'activité: premiers étayages pour une clinique de l'usage, Psychologie du Travail et des Organisations 22(1), pp. 4-21 (2016).
7. Pratt, S.R., Heintzelman, A.T., and Deming, S.E.: The Efficacy of Using the IBM Speech Viewer Vowel Accuracy Module to Treat Young Children With Hearing Impairment, Journal of Speech, Language, and Hearing Research 36(5), pp. 1063-1074, (1993).

8. Rodríguez, W.R., Saz, O., and Lleida, E.: A prelingual tool for the education of altered voices 54(5), pp. 583-600, Speech Communication (2012).
9. Navarro-Newball, A.A., Loaiza, D., Oviedo, C., Castillo, A., Portilla, A., Linares, D., and Álvarez, G.: Talking to Teo: Video game supported speech therapy, Entertainment Computing 5(4), pp. 401-412 (2014).
10. Gluck, F., Chassot, S., Liatti, S., Domingos, A., Polla, M., Valentini, G., Deriaz, M., Bétrancourt, M., and Perez, A.: FunSpeech: promoting speech production in young children with hearing disabilities', HE-Arc, HES-SO, (2019).
11. Hirsh-Pasek, K., Zosh, J.M., Golinkoff, R.M., Gray, J.H., Robb, M.B., and Kaufman, J.: Putting education in "educational" apps: Lessons from the science of learning, Psychological science in the public interest 16(1), pp. 3-34 (2015).
12. Shoukry, L., Sturm, C., and Galal-Edeen, G.H.: Pre-MEGa: A proposed framework for the design and evaluation of preschoolers' mobile educational games, in Book Pre-MEGa: A proposed framework for the design and evaluation of preschoolers' mobile educational games', pp. 385-390, (2015).
13. Robert, C.: L'utilisation de protocoles individuels expérimentaux et quasi-expérimentaux en psychologie: aspects théoriques et méthodologiques, L'Année psychologique 119, (1), pp. 55-96, (2019).
14. Parisse, C., and Le Normand, M.-T.: Une méthode pour évaluer la production du langage spontané chez l'enfant de 2 à 4 ans, Glossa (97), pp. 20-41, (2006).

Une hybridation repensée : définition, modèle, méthode

Benoît Martinet[1], Alexis Lebis[2], Mathieu Vermeulen[3], Nadine Mandran[4], Nour El Mawas [5]

[1] Université de Lorraine, CREM
benoit.martinet@univ.lorraine.fr
[2] IMT Nord Europe, Université de Lille, Centre for Digital System
alexis.lebis@imt-nord-europe.fr
[3] IMT Nord Europe, Université de Lille, Centre for Digital System
mathieu.vermeulen@imt-nord-europe.fr
[4] Université Grenoble Alpes, LIG
nadine.mandran@univ-grenoble-alpes.fr
5 Université de Lorraine, CREM
nour.el-mawas@univ-lorraine.fr

Résumé. Cet article s'intéresse à l'hybridation. Il propose une définition, un modèle et une méthode pour la conception d'unités d'apprentissages – dans des contextes de formation qui mêlent présence et distance. À partir de ces contributions, nous souhaitons développer un *plugin* pour le *Learning Management System* MOODLE. L'objectif est d'aider les enseignants face à la complexité du rôle de concepteur (de dispositifs pédago-numériques), qu'ils sont désormais amenés à endosser. Il s'agit que les enseignants contribuent à ce que notre système de formation puisse faire face aux défis pédagogiques, économiques et sociétaux qui s'imposent. Une fois introduit le contexte et les enjeux, nous examinons successivement, deux questions de recherche principales, trois états de l'art, une proposition de modèle et de méthode de conception et les études qui les valident. Nous abordons ensuite la perspective du *plugin* à travers une formalisation UML de la structure d'une formation, sur laquelle se base la méthode (HYRIS) qui est ensuite mise en œuvre. Enfin, nous concluons sur la perspective de sous-modèles qu'il reste à développer – dans le cadre d'une automatisation complète du processus de conception – et des études qu'il sera nécessaire de mener.

Mots-clés : hybridation, dispositif hybride, formation mixte, *blended learning,*

Abstract. This article is about hybridization. It provides a definition, model and method for the design of learning units – in training contexts that combine presence and distance. Based on these contributions, we want to develop a plugin for the MOODLE Learning Management System. The objective is to help teachers face the complexity of the role of designer (pedo-digital learning program), which they are now being forced to assume. Teachers are needed to help our training system meet the educational, economic and societal challenges that arise. Once the context and issues are introduced, we examine briefly two main research questions, three state of the art, a design model and method proposal and the studies that validate them. We then approach the plugin perspective through a UML formalization of the structure of a training, on which is based the method (HYRIS) that is subsequently implemented. Finally, we conclude on the perspective of sub-models that remain to be developed – as part of a complete automation of the design process – and studies that will be necessary.

Keywords: hybridization, hybrid device, mixed training, blended learning.

1 Introduction

Le système de formation français est en pleine mutation socio-économique et dans ce cadre il doit faire face à plusieurs enjeux. Des enjeux pédagogiques, car si l'on note pour l'année scolaire 2022-2023[1] une très légère baisse des inscriptions (i.e. 1,5%) dans l'enseignement supérieur, notre système de formation a toutefois accueilli ces dernières années, à chaque nouvelle rentrée, près de trois millions d'étudiants [1]. Paquelin et sa collaboratrice [2] recourent d'ailleurs à ce sujet au syntagme de « massification de l'enseignement ». Ces étudiants ont des profils hétérogènes et incluent des internationaux dans le cadre du processus initié par le traité de Bologne [3]. Ainsi, tous requièrent des réponses de formation adaptées à leurs besoins spécifiques. À ces enjeux pédagogiques, s'adjoignent des enjeux économiques et sociétaux. Des enjeux économiques, car notre pays doit rester compétitif face à une concurrence acharnée [4]. Dans ce cadre, il est essentiel de pouvoir maintenir un niveau de qualification des individus en adéquation avec les besoins économiques de notre pays. Ainsi, notre système de formation doit entre-autres permettre de se former tout au long de la vie. Enfin, il existe des enjeux sociétaux comme celui de faire de la France une société apprenante [5]. Celui-ci est non seulement relatif à la formation et au développement personnel des citoyens de demain, mais concerne également l'inclusion des publics empêchés (e.g. personnes en situation de handicap, sportifs de haut niveau, en détention [6], [7], [8]. Ainsi, selon Charlier [6], l'un des défis majeurs pour notre système de formation est d'être en mesure de proposer des dispositifs de formation, capables d'optimiser les enseignements et les apprentissages en considérant diverses contraintes.

Pour relever ces enjeux, l'institution et les chercheurs [9], [10], [11], [12] préconisent depuis plusieurs années un renouvellement des pratiques pédagogiques à partir de modalités, soutenues par les Technologies Numériques de l'Information et de la Communication (TNIC). Parmi ces modalités, Docq et ses partenaires [13] considèrent celle qui mobilise à la fois des périodes en présence et à distance, baptisée « hybride », comme particulièrement adaptée à un contexte évolutif, et Deschryver [14] comme en mesure de satisfaire les besoins spécifiques des étudiants. Djebara [15] ajoute que l'hybridation est de nature à favoriser leur réussite. Ainsi, les formations qui s'inscrivent dans ce cadre seraient plus accessibles notamment en étant plus flexibles, car en satisfaisant plusieurs types de contraintes (e.g. géographique, temporelle) [13], et également plus adaptées en mobilisant entre-autres, des pédagogies enrichies (e.g. pédagogies actives). Cependant, la modalité hybride implique notamment, des enseignements qui sont répartis sur plusieurs unités spatiales et temporelles et qui sont davantage centrés sur les apprentissages [2], [16], alors que le modèle de formation actuel est encore majoritairement organisé à partir d'une seule unité de temps et de lieu et principalement axé sur la transmission de contenus [17]. De surcroît, lorsque les phases s'exercent à distance il existe une rupture entre les enseignements et les apprentissages [18] qu'il est nécessaire de prévoir, d'organiser et de planifier [4] sur les plans symbolique, relationnel, technique et matériel [19]. Ainsi, le changement du modèle de formation qu'implique l'hybridité [20] induit de nombreuses remises en cause. Parmi-celles-ci, il est nécessaire que les étudiants développent leur autonomie et leur niveau de maîtrise technologique [21] et que les enseignants endossent de nouveaux rôles, comme celui

[1] https://www.enseignementsup-recherche.gouv.fr/fr/baisse-des-effectifs-inscrits-dans-l-enseignement-superieur-en-2022-2023-91918

de tuteur [22] ou de concepteur de dispositifs pédago-numériques, ce qui leur pose de nombreuses difficultés [23].

2 Questions et méthode de conduite de la recherche

Dans ce contexte et dans le cadre de ce travail doctoral, nous avons souhaité accompagner la conception de dispositifs pédagogiques dans des contextes qui mêlent présence et distance, ce qui suppose de répondre à deux questions de recherche principales. D'une part, « **Qu'est-ce que l'hybridation ? »** C'est-à-dire comment définir l'hybridation et la caractériser ? Et d'autre part, « C**omment concevoir et mettre en œuvre des dispositifs hybrides ? »**

Comme toute recherche, il est nécessaire de recourir à une méthode pour la conduire, celle-ci devant être en adéquation avec la posture épistémologique. Notre posture épistémologique est celle du constructivisme pragmatique [24] dans ce cadre nous avons choisi la méthode de conduite de la recherche *THEDRE* [25] car elle satisfait les contraintes de notre terrain d'études comme l'intégration de l'humain, ou encore le développement parallèle d'une connaissance et d'un outil.

3 État de l'art

À cette fin, nous avons interrogé la littérature à trois niveaux. Au niveau sociétal pour identifier les enjeux et les besoins. Au niveau académique pour connaître l'existant, et enfin, au niveau technique pour identifier les méthodes de conception existantes et les bonnes pratiques à mettre en œuvre. Nous rapportons dans les lignes qui suivent les principaux résultats de ces 3 états de l'art [26].

Au niveau sociétal, à l'aune des enjeux pédagogiques, économiques et sociétaux que synthétise Béjean [5] il y a déjà quelques années, il semble effectivement nécessaire que le système de formation fasse évoluer le modèle sur lequel il repose, afin de rendre les formations plus accessibles. Or, la continuité pédagogique instaurée par le gouvernement [27] pendant la pandémie de la COVID-19 a révélé que notre institution souffrait dans ce cadre de plusieurs faiblesses. Ainsi, sur le plan pratique l'état de l'art révèle une ingénierie artisanale [28], aboutissant le plus souvent à une transposition du présentiel à distance, plutôt qu'à une réelle transition des pratiques pédagogiques [29]. Cela a notamment donné lieu à de longues heures de visioconférence et des cimetières de PDF, traduisant la difficulté des enseignants à concevoir des dispositifs techno-pédagogiques, ainsi qu'un déficit de compétences, de formation [29], de méthodes et d'outils [30]. Par ailleurs, l'état de l'art académique nous a permis d'identifier 12 caractéristiques plus ou moins présentes (i.e articulation présence/distance, médiatisation, accompagnement, ouverture, médiation, communication, dispositifs techno-pédagogiques, autonomie, évaluation, structuration des contenus, motivation, mixes d'approches) [31] et trois points d'entrée pour l'hybridation (i.e. parcours de formation, scénario pédagogique, contenus didactiques) [32]. Cet état de l'art a également permis de se rendre compte que tout le monde parle de l'hybridation sans forcément parler de la même chose, car il existe différentes acceptions de l'hybridation (i.e. *blended,* hybride, *hybrid*, mixte). Dans ce contexte, Peltier et Séguin, [33] ont d'ailleurs proposé

une catégorisation des définitions de l'hybridation en fonction de l'axe qu'elles priorisent, comme l'ingénierie ou les approches pédagogiques. Parmi ces acceptions, les vocables les plus utilisés sont sans conteste « l'hybride » et le « *blended* ». Ainsi, les dispositifs « hybrides » au sein des publications francophones, relatent généralement une combinaison d'activités offertes en présence et à distance (selon différentes temporalités) qui reposent sur des environnements techno-pédagogiques [20], [34], [35]. Quant au *blended*, il décrit des dispositifs structurés à partir d'approches multiples et combinées de façon harmonieuse : « *A more harmonious mixing as opposed to a combination of incongruent methods* » [36] cités par Helms [37, p. 805] .
On remarque dans le corpus constitué qu'aucune proposition considère d'une part, l'ensemble des caractéristiques et les trois points d'entrée de l'hybridation et d'autre part, le respect de certains principes comme l'engagement des étudiants, l'importance de fournir des rétroactions, le développement de l'autonomie des étudiants ou encore la nécessité d'assurer la cohérence de l'articulation des périodes de formation. Par ailleurs, les méthodes existantes se révèlent être trop complexes ou inadaptées [30]. Il en ressort que l'hybridation n'est pas assez opérationnelle et qu'il est nécessaire d'accompagner les enseignants notamment, à la conception de dispositifs pédagogiques hybrides. À dessein, il est essentiel de disposer d'un modèle de conception spécifique à l'hybridation.
Enfin, l'état de l'art technique révèle : 1) l'importance de l'ingénierie pour exploiter le potentiel des nouvelles modalités et des technologies [16], [17], [38], 2) le fait qu'il est essentiel de considérer certains principes comme la qualité des interactions, l'engagement ou encore les rétroactions [20], 3) que les méthodes et les outils existants sont parfois trop complexes ou inadaptés [30].
Face à ces manques et besoins, nous avons proposé un modèle et une méthode de conception, lesquels reposent sur une proposition de définition pour l'hybridation. Celle-ci s'appuie sur 6 paramètres (i.e. articulation des activités, assistance éducative, mise en média, ouverture, évaluation, motivation) dont celui de la motivation est transversal aux 5 premiers. Ces paramètres ont été proposés sur la base des caractéristiques de l'état de l'art [31] en un nombre réduit, ce qui la rend de notre point de vue plus opérationnelle. Cette contribution, nous permet de répondre à notre première question de recherche : qu'est-ce que l'hybridation ? (i.e comment définir et caractériser l'hybridation ?).

« L'hybridation est une combinaison pédagogique ***cohérente****, de périodes de formation en co-présence physique de l'enseignant et des étudiants et de périodes à distance synchrone ou asynchrone soutenue par les technologies. Elle vise à accroître l'accessibilité des formations en les rendant à la fois plus flexibles et plus adaptées […] ».* [31].

La cohérence évoquée dans la définition est à comprendre comme l'absence de contradiction dans l'enchaînement des parties d'un tout. Elle est mise en œuvre dans nos propositions par la **continuité**, à laquelle nous avons donné un autre sens que celui qui la définissait comme un lien pédagogique [27]. À cette fin nous avons mobilisé le continuum de deux dimensions [13, p. 49], [39] dont l'une est liée au rapport présence-distance et l'autre au rapport enseigner-apprendre, et que nous avons modélisé. La continuité s'opérationnalise dans nos propositions en deux opérations dont l'ordre est interchangeable : 1) assurer la cohérence des intentions pédagogiques sous la forme de

phases qui se complètent ou se préparent, 2) identifier pour chaque intention pédagogique, la modalité spatio-temporelle la plus appropriée (i.e. présentiel[2], distanciel synchrone, distanciel asynchrone) au regard du besoin pédagogique, du contexte et des contraintes. La combinaison de ces phases et spatio-temporalités permettant de qualifier le dispositif dans sa globalité (e.g. hybride, hybride synchrone, comodal).

4 Un modèle et une méthode pour concevoir des dispositifs hybrides

4.1 Un modèle de conception

Nous avons proposé un modèle de conception que nous avons baptisé « *Grid for Reading HYbridization Device* » (*THE GRHYD*). Il considère la définition et les paramètres de l'hybridation, et permet de mettre en œuvre la continuité.

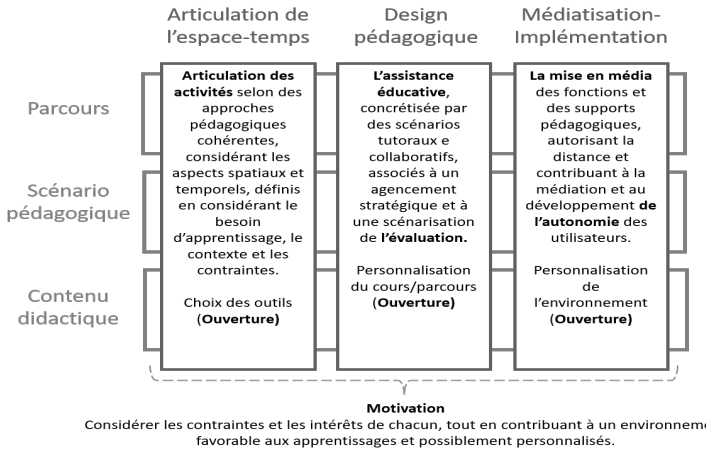


Figure 1: Structure du modèle de conception

Sa structure (*Figure 1*) repose sur la combinaison des 3 points d'entrée pour l'hybridation et de 3 concepts théoriques (e.g. l'apprentissage instrumenté par les scénarios [40], des catégories de composantes propres à tout dispositif [41], des méthodes de conception « ADDIE » [42] et « MISA » [43]) qui n'avaient jamais été articulés de cette manière. Le modèle est ainsi constitué de 3 niveaux et de 3 phases. Le niveau « Macro » noté (M) correspond à de l'ingénierie de formation. Il aboutit à un parcours de formation médiatisé. Il intéresse les ingénieurs de formation, les ingénieurs pédagogiques, les directions d'établissements et les enseignants. Il est particulièrement intéressant dans le cadre de la mobilité des étudiants. Le niveau « méso » noté (m) correspond à de l'ingénierie pédagogique, il aboutit à la réalisation de scénarios de séances / séquences. Il vise à optimiser l'impact des enseignements dans des contextes hybrides. Il intéresse

[2] Nous considérons dans nos travaux d'une part, que des activités différentes réalisées par des étudiants dans un même lieu et un même temps relèvent de la différenciation pédagogique et non pas de l'asynchronisme et d'autre part, que lorsque l'enseignant n'est pas en coprésence physique des étudiants ou d'une partie des étudiants il s'agit de distanciel.

plus particulièrement les enseignants et les ingénieurs pédagogiques. Le niveau micro noté (µ) correspond à de l'ingénierie didactique. Il vise la conception de supports (e.g. activité, ressource) et aboutit à leur structuration et médiatisation. Il intéresse les enseignants et les experts disciplinaires et permet de considérer les préférences d'apprentissage et l'inclusion. Les phases sont quant à elle le fruit d'une combinaison entre 3 catégories de composantes communes à tout dispositif [41] et des étapes propres à certaines méthodes de conception, dont la première permet de considérer la continuité.
Les intersections entre les phases et les niveaux correspondent à des étapes. Elles sont au nombre de 9 (*Figure 2*). Chaque phase est influencée par celle qui la précède et chaque niveau par celui qui lui est supérieur. Les choix sont donc raffinés à chaque niveau. Toutefois chaque niveau est également un point d'entrée pour hybrider un dispositif. Dans ce cadre il est parfois nécessaire de faire des choix, qui sont pourtant relatifs à un niveau supérieur.

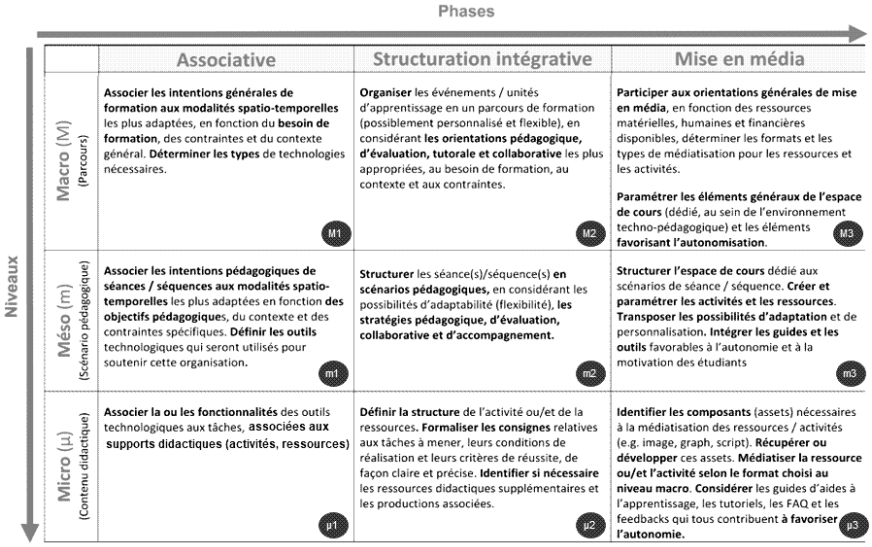


Figure 2 : The GRHYD, un modèle pour la conception de dispositifs hybrides

Les utilisateurs qui recourent à ce modèle de conception peuvent ainsi l'envisager comme un outil de cadrage et d'analyse pour se familiariser avec la conception de dispositifs hybrides, mais également comme un modèle d'ingénierie et de réingénierie lorsqu'ils recourent aux 9 guides, dont chacun correspond à une étape. Chaque guide est constitué de deux activités dont l'une est associée à une phase réflexive sous la forme de questions à se poser et, l'autre, à une phase de production qui permet de mettre en œuvre la première. Si ce modèle de conception représente un intérêt pour les enseignants, les études que nous avons menées [26] révèlent que le niveau qui intéresse le plus les enseignants est le niveau « méso ». C'est la raison pour laquelle nous avons proposé une méthode instanciée à partir de ce niveau.

4.2 Une méthode de conception

Nous avons baptisé cette méthode *HYbRid Innovative Scenari* (HYRIS). Cette méthode est accompagnée d'un guide, lequel assiste les enseignants dans la réalisation de trois tâches qu'ils doivent mener.

- La première tâche, celle de design, permet de faire des choix quant au design hybride d'une Unité d'Apprentissage (UA). Une UA correspond à une unité de structuration qui vise l'atteinte d'un objectif général ou d'une compétence et qui peut se dérouler sur une ou plusieurs séances (i.e. séquence). Le « design hybride » s'opère notamment, à l'aide d'une roue de l'hybridation (*Figure 3*). Cette tâche intègre une phase de cadrage visant entre autres à déterminer le besoin et les contraintes. La roue est interactive et permet de faire des choix par rapport aux paramètres et sous paramètres de l'hybridation et fournit des informations pour les mettre en œuvre.

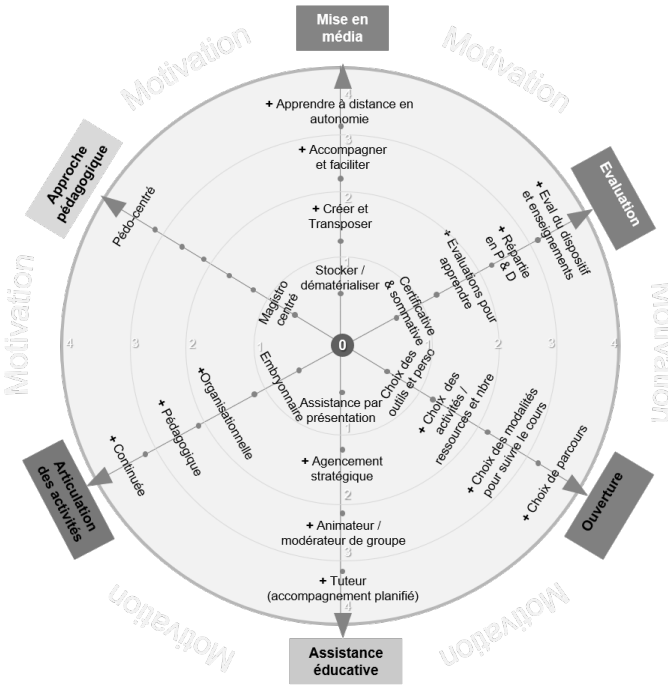


Figure 3 : Roue de l'hybridation définie à partir de notre proposition de caractérisation de l'état de l'art

- La seconde tâche est celle de méta-scénarisation. Celle-ci vise l'articulation spatio-temporelle de périodes à partir d'intentions pédagogiques définies sous la forme de phases, lesquelles préstructurent l'UA en se complétant ou / et se préparant mutuellement. Les phases se décomposent en objectifs spécifiques, lesquels sont caractérisés à partir de variables relatives aux modèles de formation (i.e. spatio-temporalité, organisation du groupe, façon d'apprendre.) Ainsi, nous suggérons que ces phases soient déterminées à partir de méta-activités – existantes ou à construire – (e.g étude de cas, approche par problème, classe inversée) et de plus, en adéquation avec l'approche par

compétences [44]. Nous définissons les méta-activités comme des activités qui ont la particularité de se décomposer en plusieurs phases, ce qui les rend particulièrement adaptées à la continuité.

- La troisième tâche est celle de scénarisation détaillée. Cette dernière poursuit le processus de scénarisation initié lors de la méta-scénarisation, en décrivant la façon dont seront atteints chacun des objectifs spécifiques. À cette fin, sont mobilisés 7 familles d'élèments [26] que nous avons définis à partir de l'état de l'art (i.e. objets, modalités spatio-temporelles, formats médiatiques, organisation du groupe, rôles de l'enseignant et assistance, règles de démarche et d'adaptabilité, les types et méthodes et techniques d'évaluation).

Plusieurs limites apparaissent cependant à cette méthode : 1) elle requiert de nouvelles évaluations suite aux améliorations apportées dans le cadre d'une première étude, 2) elle ne dispose pas d'un outil qui catalyse les 3 tâches à mener, 3) l'exhaustivité souhaitée pour représenter l'ensemble des dispositifs peut la rendre complexe et requérir que les enseignants soient assistés.

5 Méthode de production des données et résultats

Plusieurs études ont été menées pour évaluer ces contributions. Toutes ont mobilisé la méthode des *focus-groups* afin d'observer un maximum d'avis et recueillir un maximum de propositions des participants. Ainsi, la méthode de production et d'analyse des données de ces études est qualitative. Elles recourent certes à une dimension quantitative à travers des échelles de Likert, mais le nombre de partcipants à ces études n'est pas suffisant pour caractériser notre expérience de quantitative. De plus, aucune méthode d'échantillonnage au sens statistique du terme ne peut être élaborée.

La première étude (GEP1) vise à évaluer le modèle pour la conception de dispositifs hybrides (lequel a été proposé en considérant notamment, une première définition de l'hybridation). Les participants sont au nombre de six, parmi lesquels deux ingénieurs pédagogiques, deux conseillers pédagogiques et deux enseignants chercheurs. Tous sont habitués à la formation à distance et ont été répartis dans les groupes de façon à ce que chaque métier soit représenté. Le *tableau 1* synthétise les questions, les objectifs et les activités mises en place.

Tableau 1. Synthèse des questions, des objectifs et des activités associées à l'étude GEP1

Question	Objectif	Activité
La définition de l'hybridation proposée est-elle acceptée par le panel d'experts ?	Évaluer si la proposition de définition de l'hybridation est en adéquation avec les représentations des participants.	-Lire la définition en individuel, en débattre en sous -groupes puis formaliser une synthèse. -Débriefing.
Les contenus qui décrivent les étapes du modèle sont-ils cohérents/appropriés ? (i.e rendre propre à un usage).	Évaluer la cohérence du contenu des étiquettes qui décrivent les étapes du modèle de conception.	-En individuel lire le contenu des étiquettes, puis placer celles-ci sur une matrice qui représente les 9 étapes du modèle de conception. - En individuel et à l'aide d'un code couleur, évaluer l'intelligibilité du contenu de chaque étiquette. -Débriefing.
Le concept d'une ingénierie pour la conception de dispositifs hybrides en trois niveaux et trois phases est-il utile/pertinent, conforme à la norme, facile à interpréter et cohérent ?	Évaluer le concept d'une ingénierie hybride à plusieurs niveaux et plusieurs états.	-En individuel, répondre à un questionnaire en ligne, où les réponses sont apportées au moyen d'une échelle de Likert et complétées si nécessaires par des commentaires. -Débriefing.

La méthode d'analyse des réponses est fonction de la façon selon laquelle ont été collectées les données. Ainsi, au sujet de la définition, nous avons analysé les synthèses produites par les groupes, sous la forme de points forts et de points faibles. Pour les deux autres questions, nous avons d'une part, analysé les réponses issues des échelles de Likert à l'aune du nombre de points obtenus que nous avons ensuite remis en perspective à l'aune des libellés associés, et d'autre part, à partir des commentaires et suggestions formulées par les participants. Le *tableau 2*, synthétise l'objet de cette étude, les résultats obtenus et les actions mises en oeuvre.

Tableau 2. Objet de l'étude GEP1, résultats et actions

Objet de l'étude	Résultats obtenus	Actions mises en oeuvre
Définition	Pas de remise en question du fond, mais est perfectible sur la forme.	Réviser la définition en la simplifiant à partir d'un état de l'art approfondi.
Contenus qui décrivent les étapes du modèle	Des contenus cohérents mais parfois trop complexes.	Clarifier et simplifier les contenus.
Des niveaux et des phases	Un cadre théorique très intéressant, mais qui reste trop complexe et pass assez opérationnel.	Associer des guides à chacune des étapes du modèle pour accompagner les utilisateurs. Associer des verbes d'action aux contenus.

La seconde étude (GEP2) vise à évaluer la pertinence et l'utilité des guides associés à chacune des 9 étapes de notre modèle de conception (*Figure 2*). Chaque guide est composé d'une activité de réflexion sous la forme de questions à se poser, et d'une activité de production qui les met en oeuvre. Seize ingénieurs pédagogiques ont été recrutés et répartis en groupe de façon aléatoire, avec pour tâche d'évaluer les guides associés à l'un des niveaux du modèle de conception (i.e. Macro, méso, micro) auquel ils sont affectés. Nous aurions aimé que les participants puissent analyser l'ensemble des 9 guides, malheureusement, pour des raisons temporelles cela n'a pas été possible. Le *tableau 3* synthétise les questions, les objectifs et les activités proposées.

Tableau 3. Synthèse des questions, des objectifs et des activités associées à l'étude GEP2

Question	Objectif	Activité
Les questions associées à chacun des guides, sont-elles utiles et pertinentes par rapport à l'étape du modèle à laquelle elles se rapportent ?	Évaluer la pertinence et l'utilité des jeux de questions.	-Lire les questions en individuelle - En débattre en sous-groupe - Évaluer les questions au moyen du questionnaire en ligne, lequel intègre la possibilité pour les participants d'ajouter des commentaires.
Les questions qui sont associées à chacun des guides, permettent-elles d'assurer la complétude des points à aborder pour l'étape du modèle à laquelle elles se rapportent ?	Évaluer si les jeux de questions assurent la complétude des points à aborder pour l'étape à laquelle ils se rapportent.	-Formaliser d'éventuelles propositions.
L'activité de production contribue-t-elle à la conception de l'étape à laquelle elle se rapporte ?	Évaluer si les activités de production associées aux guides contribuent à la conception de l'étape à laquelle ils se rapportent.	-En individuel, répondre à un questionnaire en ligne, où les réponses sont apportées au moyen d'une échelle de Likert et complétées si nécessaires par des commentaires. -Débriefing

Les données recueillies sont sous deux formes : des questions ouvertes pour obtenir des suggestions et des échelles de Likert, dont nous avons analysé les données de façon identique à la première étude. Les questionnaires relatifs à l'évaluation des activités de production, ont été élaborés à partir d'une adaptation des catégories d'indicateurs de qualité de données de Ruocco (2012). Ces questionnaires, ont pour objet de mesurer l'utilité (i.e. dans le sens de l'intérêt), la conformité aux standards habituels, la facilité d'interprétation et la cohérence. Quant aux questionnaires d'évaluation, relatifs aux jeux de questions, ils visent à évaluer la pertinence (i.e. dans le sens d'être approprié) et l'utilité de chaque question. Les questionnaires ont été conçus par niveau, sous la forme d'un parcours d'évaluation. Nous avons utilisé à cette fin l'outil « Limesurvey ». Le *tableau 4* synthétise l'objet de l'étude, les résultats obtenus et les actions mises on oeuvre.

Tableau 4. Objet de l'étude, résultats et actions de l'étude GEP2

Objet de l'étude	Résultats obtenus	Actions mises en oeuvre
Utilité et pertinence des jeux de questions	Au niveau macro, certains éléments ne semblent pas toujours faire sens. Au niveau méso la continuité ne semble pas aller de soi. Au niveau micro les résultats sont satisfaisants, mais à mettre en perspective du faible nombre de participants.	Sur le fond ajouter des questions, clarifier certaines notions, et exemplifier. Sur la forme ancrer de nouvelles considérations (e.g. paramètres de l'hybridation, continuité).
Contribution des activités de production à la conception de dispositifs	Au niveau macro, de très bons résultats mais l'hybridation des parcours semblent plus difficiles à mettre en perspective. Au niveau méso, les résultats sont plus nuancés. La continuité et le niveau de détail de la scénarisation, semblent ne pas être habituel chez les participants. Au niveau micro, les propositions contribuent assez bien à la conception de dispositifs.	Sur le fond, clarifier certaines notions et harmoniser les terminologies. Sur la forme, recentrer les propositions sur le niveau méso et accompagner les enseignants face à un modèle qui reste complexe.

La troisième étude (NGE) vise à évaluer la méthode de conception HYRIS. Il s'agit d'un part, d'évaluer la tâche de design (les 2 autres tâches ayant déjà été évaluées à

travers les études GEP-1 et GEP-2) et d'autre part, d'évaluer le guide d'accompagnement. Le *tableau 5* synthétise les questions, les objectifs et les activités proposées. Le panel de participants est composé de huit praticiens (issus d'écoles d'ingénieurs de l'Institut Mines Telecom). Le recrutement a eu lieu en deux temps. Le premier a consisté à identifier des ingénieurs pédagogiques volontaires au sein de ces écoles, le second à charger ceux-là de recruter des enseignants chercheurs intéressés.

Tableau 5. Synthèse des questions, des objectifs et des activités associées à l'étude NGE.

Question	Objectif	Activité
Le guide permet-il d'accompagner les participants à travers les tâches qu'ils doivent mener ?	-Évaluer l'utilité, la cohérence et la complétude du guide d'accompagnement.	-Utiliser le guide pour réaliser les activités proposées. -Remplir le questionnaire adapté de Di-Ruocco.
Le guide est-il utilisable ?	-Évaluer si le guide est utilisable.	-Remplir le questionnaire *SUS* de Brooke.
Aucune question associée	-Informer les participants.	Lire en individuel les notions importantes relatives à l'hybridation.
La tâche de design permet-elle d'opérationnaliser le design hybride de l'unité d'apprentissage ?	-Évaluer si l'activité relative au design hybride de l'unité d'apprentissage est utile, cohérente, complète ?	*En binôme (enseignant/ingénieur)* -Formaliser les caractéristiques et la description de l'unité d'apprentissage. -Identifier les orientations spatio-temporelles de l'unité d'apprentissage -Sélectionner les niveaux pour les paramètres de l'hybridation. - Identifier comment les mettre en œuvre.
L'activité relative au design hybride de l'unité d'apprentissage est-elle utilisable ?	-Évaluer l'utilisabilité de l'activité de design.	-Remplir le questionnaire adapté à partir de Di-Ruocco. -Remplir le questionnaire *SUS* de Brooke.

Hormis les questions ouvertes destinées à recueillir des commentaires et des suggestions, les autres données ont été collectées à partir d'échelles de Likert. L'une d'entre elles vise à évaluer l'utilité, la cohérence et la complétude du guide et de la tâche de design, l'autre, a pour but évaluer leur utilisabilité. La première échelle a été construite à partir d'une adaptation des critères de qualité des données de Ruocco (2012), la seconde en mobilisant l'échelle *SUS* de Brooke [46]. Le *tableau 6* synthétise l'objet de l'étude, les résultats obtenus et les actions mises on oeuvre.

Tableau 6. Objet de l'étude, résultats et actions de l'étude NGE

Objet de l'étude	Résultats obtenus	Actions mises en oeuvre
Guide d'accompagnement	Résultats très satisafisants du point de vu de la coherence, de l'utilité et de la complétude et très encourageants au niveau de l'utilisabilité.	Homogénéiser les titres entre les différents supports. Compléter l'étude d'utilisabilité par des entretiens.
Tâche de Design hybride	Résultats très satisfaisants du point de vu de la cohérence, de l'utilité et de la complétude et très moyennement satisfaisants au niveau de l'utilisabilité.	Utiliser des termes moins complexes, ajouter une slide relative à l'articulation des activités, modifier 2 échelles de la roue de l'hybridation. Faire figurer les niveaux d'échelles sur les ressources complémentaires.

6 Vers un plugin MOODLE pour la conception d'unités d'apprentissage

Afin de surmonter la difficulté liée à la richesse de notre proposition et pour limiter le temps investi par les enseignants pour concevoir une UA hybride dans le *LMS* MOODLE (le plus répandu au sein des universités françaises[3]), nous souhaitons proposer un *plugin* d'asssistance.

6.1 Assistance et modalités

Nous envisageons cette assistance selon deux modalités. La première modalité consiste à faciliter la conception à partir de gabarits, la seconde à assister la conception à partir « de zéro ». Nous entendons par gabarits, des templates, lesquels préstructurent notamment, de façon pédagogique et spatio-temporelle une UA. Nous les envisageons comme évoqué à partir de méta-activités. La deuxième modalité, celle qui est relative à la conception d'une UA à partir de zéro, procède en un processus en 6 étapes (Figure 4) qui suit la méthode HYRIS (i.e. cadrage, design hybride, options de mise en œuvre de l'hybridation, méta-scénarisation, scénarisation détaillée).

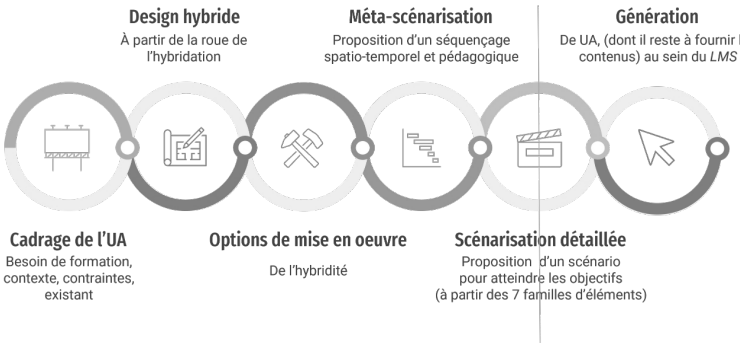


Fig. 4. Processus d'assistance à la conception d'une UA pour LMS à partir de zéro

La première étape est relative au cadrage de l'UA. L'enseignant concepteur est amené à saisir des informations qui permettent de caractériser l'UA, identifier le contexte et

[3] https://www.elearningtouch.com/serie-idees-recues-plateforme-lms-moodle/ (publié et consulté le 23 juin 2024).

les contraintes (e.g. libellé, description, volume horaire, objectif général / compétence, nombre de séances, nombre d'étudiants, charge de travail, contexte et contraintes pédagogiques, matérielles et techniques). La seconde étape vise le design hybride de l'UA, au moyen de la roue de l'hybridation. La troisième étape, au regard des niveaux de paramètres choisis pour chaque axe de la roue, s'attache à sélectionner des options qui permettent de les mettre en œuvre. La quatrième étape consiste en la validation ou la modification d'une méta-scénarisation, proposée de façon automatique. La cinquième étape consiste également en une proposition, qu'il faut confirmer ou modifier. Celle-ci est relative à la scénarisation détaillée, et est notamment, proposée sur la base de la méta-scénarisation et du niveau taxonomique des sous-objectifs de phases (objectifs spécifiques). La sixième étape consiste à générer l'UA à partir d'un dernier clic de souris. Il est à noter que l'UA générée reste totalement muable, soit en amendant directement l'un des éléments au sein de la plateforme ou encore au moyen du tableau de bord et de l'interface d'assistance, qui restent mobilisables à tout instant.

Pour réaliser ce *plugin* nous proposons de formaliser les relations qui s'exercent entre les éléments qui structurent le dispositif hybride lors de la tâche de méta-scénarisation. Nous souhaitons ainsi, fournir à la communauté un socle théorique et technique nécessaire pour envisager la mise en place d'assistance pour les enseignants.

6.2 Proposition de modélisation et implications

Le point important du *plugin* que nous proposons concerne la structuration de(s) UA à partir de la tâche de méta-scénarisation et doit donc en constituer la base. Le schéma de classe UML (*Figure 5*), en permet une modélisation (classique pour la conception en informatique).

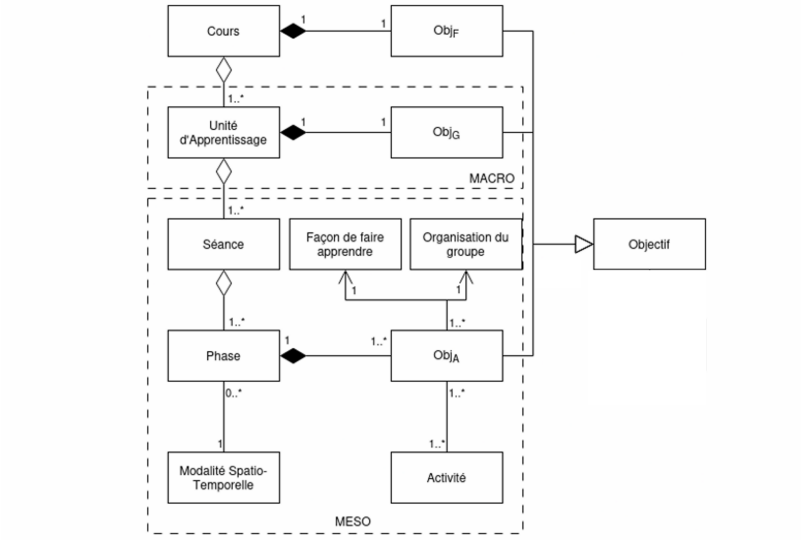


Fig. 5. Modélisation UML de la structuration de(s) UA à partir de la tâche de méta-scénarisation de la méthode HYRIS (ObjF, ObjG, ObjA signifient respectivement Objectif de Formation, Objectif Général et Objectif d'Apprentissage).

Ainsi, notre proposition permet également de suggérer aux utilisateurs des configurations de formations basées sur des élèments du modèle récurrents et plebiscités par les enseignants. Par exemple, en décrivant un objectif d'apprentissage $o \in ObjA$, *avec* Obj_A l'ensemble des objectifs d'apprentissage, comme constitué d'un ensemble d'activités $\alpha_o \subseteq \Omega_\alpha$, avec Ω_α l'ensemble des activités décrites par les enseignants, nous pouvons recouper l'ensemble des activités communes α_c d'un ensemble d'objectifs d'apprentissage $\mathcal{O} \subseteq Obj_A$, tel que $\alpha_c = \bigcap_{i=1}^{card(\mathcal{O})} \alpha_{o,i}$ et ainsi permettre aux enseignants d'identifier des activités pertinentes pour mobiliser des compétences souhaitées. La recommandation d'activités pour un nouvel objectif d'apprentissage est également possible par ce biais en définissant une fonction subjective $\mu : Obj_A \rightarrow Obj_A$, tel que l'image de o par μ est l'ensemble des objectifs d'apprentissage les plus similaires à o.

Enfin, notre modèle permet la qualification automatique du type d'hybridation des phases, des séances, des unités d'apprentissage et des cours en combinant les modalités spatio-temporelles $\mathcal{M} = \{p, d_s, d_a\}$, à savoir présentiel ($p$), distanciel synchrone (d_s) et distanciel asynchrone (d_a), de chacune des phases. Un élément du modèle est ainsi qualifié d'hybride dans notre modèle si, et seulement si, au moins deux types de modalités sont présentes. Nous formalisons donc la notion d'hybridation comme suit:

- Hybride $H_1 = \{p, d_a\}$, - Hybride distant $H_2 = \{d_a, d_s\}$,
- Hybride synchrone $H_3 = \{p, d_s\}$, - Co-modal $H_4 = \{p, d_s, d_a\}$

Bien que nous envisageons de développer un *plugin* pour MOODLE sur la base de ce modèle UML, et puisque ce dernier est générique, il sera possible d'en développer pour d'autres plateformes.

7 Conclusion et perspectives

Sur la base des revues de littérature nous avons dans cet article identifié plusieurs acceptions de l'hybridation et des paramètres ainsi, que des cadres théoriques, lesquels nous ont permis de proposer trois contributions validées par les études : 1) une définition de l'hybridation, 2) un modèle et 3) une méthode pour la conception de dispositifs hybrides. Il est à noter que parmi ces acceptions, de notre point de vue, l'hybridation repose avant tout sur une articulation cohérente de périodes de formation en présence et à distance, (qui sont menées de façon synchrone ou / et asynchrone). Pour autant, cela n'est pas en contradiction avec la possibilité d'intégrer diverses approches pédagogiques de nature à faciliter les enseignements et les apprentissages, choisies et combinées en fonction du besoin, du contexte et des contraintes.

Le modèle proposé est adapté à la conception de dispositifs hybrides premièrement, car il intègre l'ensemble des paramètres de l'hybridation et deuxièmement, car il répond à trois besoins majeurs de la communauté éducative (i.e. hybrider des parcours de formation, hybrider des scénarios pédagogiques, proposer des contenus didactiques adaptés aux différents publics) [32]. Ainsi, si le modèle de conception est sans doute le plus utile aux établissements, aux ingénieurs de formation et aux ingénieurs pédagogiques, il est également très intéressant pour les enseignants. La méthode de conception que nous proposons met l'accent sur le niveau pédagogique du modèle (i.e. niveau méso). Elle répond ainsi à la préoccupation majeure des enseignants ; celle de concevoir des séances et des séquences hybrides. Cependant, l'exhaustivité des éléments que mobilise

la méthode, – dans le souci d'avoir la capacité de représenter l'ensemble des dispositifs existants ou à construire – requiert le plus souvent que les enseignants soient assistés par un ingénieur pédagogique. Pour surmonter cette difficulté sans pour autant dégrader l'exhaustivité de la méthode, nous proposons de développer à partir de celle-ci, un *plugin* d'assistance pour le *LMS* MOODLE dont nous avons initié la modélisation *UML* et les implications.

Nous espérons que ces contributions permettront de mieux accompagner les enseignants, afin de leur éviter de simplement transposer des pratiques habituelles en situation de distanciel, qui conduisent bien souvent à de longues heures de visioconférences ainsi qu'à des « cimetières » de PDF. Cependant, puisque nous visons une automatisation de la conception à partir d'un *plugin*, plusieurs travaux restent à mener. Nous devrons notamment, développer des sous-modèles (associés à certains des paramètres de notre proposition). Parmi ceux-là, l'un est au sujet de l'accompagnement des étudiants. La problématique étant alors d'identifier entre-autres, la nature (e.g. cognitif, socio-affectif), le type (e.g. proactif, réactif), les protagonistes, le moment et les outils qui doivent être apportés aux étudiants, de façon qu'ils puissent accomplir les activités d'apprentissage. Par ailleurs, nous devrons mener plusieurs études : 1) pour évaluer les sous-modèles, 2) le *plugin (e*.g. utilisabilité, utilité), 3) et enfin pour mesurer la pertinence et l'impact des UA qui ont été générées. Toutefois, les perspectives qu'offre notre travail sur la mutualisation des pratiques pédagogiques nous semblent une des perspectives les plus importantes.

Références

[1] MESR, « Prévisions des effectifs dans l'enseignement supérieur - rentrées 2021 et 2022 », enseignementsup-recherche.gouv.fr. Consulté le: 11 mars 2024. [En ligne]. Disponible sur: https://www.enseignementsup-recherche.gouv.fr/fr/previsions-des-effectifs-dans-l-enseignement-superieur-rentrees-2021-et-2022-50825

[2] D. Paquelin et L. Lachapelle-Bégin, « Hybridation: principes et repères », Université de Laval, Québec, Rapport de recherche, 2022. [En ligne]. Disponible sur: https://hal.science/hal-03718900/document

[3] Ministère Enseignement supérieur et Recherche, « Présentation de l'EEES », enseignementsup-recherche.gouv.fr. Consulté le: 4 avril 2024. [En ligne]. Disponible sur: https://www.enseignementsup-recherche.gouv.fr/fr/presentation-de-l-eees-46573

[4] D. Peraya et C. Peltier, « Ingénierie pédagogique : vingt fois sur le métier remettons notre ouvrage… », *Distances et médiations des savoirs. Distance and Mediation of Knowledge*, nº 29, Art. nº 29, mars 2020, Consulté le: 17 juin 2021. [En ligne]. Disponible sur: http://journals.openedition.org/dms/4817

[5] S. Béjean, « Pour une société apprenante - Propositions pour une stratégie nationale de l'enseignement supérieur », Ministère de l'Education nationale, de l'Enseignement supérieur et de la Recherche, Synthèse de rapport, sept. 2015. Consulté le: 3 octobre 2023. [En ligne]. Disponible sur: https://www.enseignementsup-recherche.gouv.fr/fr/pour-une-societe-apprenante-propositions-pour-une-strategie-nationale-de-l-enseignement-superieur-47757

[6] B. Charlier, « Colloque national Hybridation des formations. Intervention grand témoin », Ministère de l'Enseignement supérieur et de la Recherche, Paris, 2023.

Consulté le: 23 janvier 2023. [En ligne]. Disponible sur: https://www.enseignementsup-recherche.gouv.fr/sites/default/files/2023-02/synth-se-du-colloque-hybridation-des-formations-par-bernadette-charlier-26433.pdf

[7] F. Henri, « Quel changement à l'ère numérique? Quelle ingénierie pédagogique pour y répondre? », *Médiations et médiatisations*, nº 2, Art. nº 2, nov. 2019, doi: 10.52358/mm.vi2.102.

[8] MENS1707714D, *Décret n° 2017-619*. 2017. Consulté le: 20 septembre 2022. [En ligne]. Disponible sur: https://www.legifrance.gouv.fr/eli/decret/2017/4/24/2017-619/jo/texte

[9] C. Bertrand, « Soutenir la transformation pédagogique dans l'enseignement supérieur », DGESIP, Paris, 2014.

[10] J.-F. Céci, « Innovation pédagogique et hybridations, au prisme du numérique : une taxonomie de l'innovation pédagogique instrumentée », présenté à Colloque International TICEMED 13 - Hybridation des formations : de la continuité à l'innovation pédagogique ?, Athènes, oct. 2022.

[11] P.-F. Coen, « L'impact réel des technologies de l'information et de la communication sur la forme scolaire », *Recherches en éducation*, nº 10, Art. nº 10, mars 2011, doi: 10.4000/ree.4755.

[12] M. Lebrun, *Théories et méthodes pédagogiques pour apprendre et enseigner. Quelle place pour les TIC dans l'éducation ?* De Boeck, 2007.

[13] F. Docq, M. Lebrun, et D. Smidts, « Analyse des effets de l'enseignement hybride à l'université : détermination de critères et d'indicateurs de valeurs ajoutées », *ritpu*, vol. 7, nº 3, p. 48-59, 2010, doi: 10.7202/1003563ar.

[14] N. Deschryver, « Interaction sociale et expérience d'apprentissage en formation hybride », Thèse de doctorat, Université de Genève, Genève, 2008. [En ligne]. Disponible sur: https://archive-ouverte.unige.ch/files/downloads/0/0/0/2/1/8/6/1/unige_21861_attachment01.pdf

[15] A. Djebara, « La pédagogie numérique : un défi pour l'enseignement supérieur », Niort, Avis du Conseil économique, social et environnemental sur le rapport présenté, 2015.

[16] E. Nissen, *Formation hybride en langues: articuler présentiel et distanciel*. in Collection "Langues & didactique". Les Editions Didier, 2019. [En ligne]. Disponible sur: https://hal.science/hal-01941203

[17] D. Peraya et J.-F. Cerisier, « Concevoir aujourd'hui des formations hybrides ou à distance. Écrit introductif au guide de l'ingénierie de formation à l'IH2EF », in *Guide de l'ingénierie de formation à l'IH2EF pour une assise scientifique de l'hybridation*, IH2EF, 2022, p. 24. Consulté le: 30 octobre 2023. [En ligne]. Disponible sur: https://archive-ouverte.unige.ch/unige:170281

[18] G. Jacquinot, « Apprivoiser la distance et supprimer l'absence ? ou les défis de la formation à distance », *rfp*, vol. 102, nº 1, p. 55-67, 1993, doi: 10.3406/rfp.1993.1305.

[19] C. Peltier, « Représentation des médias et appropriation des dispositifs médiatiques chez des enseignants du supérieur », Thèse de doctorat, University of Geneva, 2016. doi: 10.13097/archive-ouverte/unige:85010.

[20] F. Lafleur et G. Samson, *Etat de situation sur l'hybridité de la formation à distance en contexte postsecondaire*. in Formation à distance Distance learning. Québec: Presses de l'université du Québec, 2020.

[21] F. Bédart, P. Pelletier, et C. Le clech, « Échange de bonnes pratiques entre enseignants de niveau universitaire. L'apprentissage hybride. », *Université du Québec*, vol. 6, nº 1, p. 2, 2017.

[22] J. Basque, « En quoi les TIC changent-elles les pratiques d'ingénierie pédagogique du professeur d'université? », *Revue internationale des technologies en pédagogie universitaire*, vol. 1, nº 3, p. 7, 2004, doi: 10.18162/ritpu.2004.52.

[23] N. El Mawas, L. Oubahssi, et P. Laforcade, « A method for making explicit LMS instructional design languages », *Technology, Instruction, Cognition and Learning (TICL) journal*, vol. 10, nº 3, p. 203-226, 2016, Consulté le: 23 novembre 2022. [En ligne]. Disponible sur: https://hal.archives-ouvertes.fr/hal-01458143

[24] M.-J. Avenier, « Paradigmes épistémologiques constructivistes : post modernisme ou pragmatisme », *Management Avenir*, vol. 43, nº 3, p. 372-391, sept. 2011, Consulté le: 17 mai 2022. [En ligne]. Disponible sur: https://www.cairn.info/revue-management-et-avenir-2011-3-page-372.htm

[25] N. Mandran, *Traceable Human Experiment Design Research: Theoretical Model and Practical Guide*. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2018. doi: 10.1002/9781119453635.

[26] B. Martinet, « Accompagner la conception de dispositifs pédagogiques dans des contextes mêlant présence et distance », Theses, Université de Lorraine, 2024. [En ligne]. Disponible sur: https://theses.hal.science/tel-04648581

[27] S. Wagnon, « La continuité pédagogique : méandres et paradoxes en temps de pandémie », *Recherches & éducations*, nº HS, Art. nº HS, mai 2020, doi: 10.4000/rechercheseducations.10451.

[28] D. Peraya et C. Peltier, « Ce que la pandémie fait à l'ingénierie pédagogique et ce que la rubrique peut en conter », *Distances et médiations des savoirs. Distance and Mediation of Knowledge*, nº 30, Art. nº 30, juin 2020, Consulté le: 17 juin 2021. [En ligne]. Disponible sur: http://journals.openedition.org/dms/5198

[29] E. Villiot-Leclercq, « Ingénierie pédagogique au temps de la Covid-19 », *Distances et médiations des savoirs. Distance and Mediation of Knowledge*, nº 30, Art. nº 30, juin 2020, Consulté le: 16 juin 2021. [En ligne]. Disponible sur: http://journals.openedition.org/dms/5203

[30] B. Martinet, A. Lebis, et M. Vermeulen, « Vers un outil d'accompagnement pour l'hybridation de situations pédagogiques », in *Questions de Pédagogie dans l'Enseignement Supérieur*, La Rochelle, France, janv. 2022. [En ligne]. Disponible sur: https://hal.archives-ouvertes.fr/hal-03583964

[31] B. Martinet, M. Vermeulen, J. Heutte, N. Mandran, et N. E. Mawas, « L' Hybridation : anatomie d'un concept polysémique », *Sciences et Technologies de l'Information et de la Communication pour l'Education et la Formation*, vol. 30, nº 2, Art. nº 2, 2023, doi: 10.23709/sticef.30.2.5.

[32] Ministère Enseignement supérieur et Recherche, *Colloque national Hybridation des formations*, (18 janvier 2023). Consulté le: 3 octobre 2023. [Vidéo you]. Disponible sur: https://www.enseignementsup-recherche.gouv.fr/fr/colloque-national-hybridation-des-formations-88723

[33] C. Peltier et C. Séguin, « Hybridation et dispositifs hybrides de formation dans l'enseignement supérieur : revue de la littérature 2012-2020 », *Distances et médiations des savoirs. Distance and Mediation of Knowledge*, nº 35, Art. nº 35, oct. 2021, doi: 10.4000/dms.6414.

[34] B. Charlier, N. Deschryver, et D. Peraya, « Apprendre en présence et à distance a la recherche des effets des dispositifs hybrides », *Distances et savoirs*, vol. Vol.

4, nº 4, p. 469-496, 2006, Consulté le: 17 décembre 2020. [En ligne]. Disponible sur: https://www.cairn.info/revue-distances-et-savoirs-2006-4-page-469.htm

[35] DNE-TN2, « Enseignement et apprentissage en situation hybride : Portfolio du GTnum ADEF #formescolairehybride », Éducation, numérique et recherche. Disponible sur: https://edunumrech.hypotheses.org/10555

[36] R. T. Osguthorpe et C. R. Graham, « Blended learning environments: Definitions and directions », *Quarterly review of distance education*, vol. 4, nº 3, Art. nº 3, 2003, Disponible sur: https://www.learntechlib.org/p/97576/?fbclid=IwAR3zr0WF08Ju5WRZINWGolQc7UkDFFtbicLfqhbfTNcRP5-QqqfzKqGtog

[37] S. A. Helms, « Blended/hybrid courses: a review of the literature and recommendations for instructional designers and educators », *Interactive Learning Environments*, vol. 22, nº 6, p. 804-810, nov. 2014, doi: 10.1080/10494820.2012.745420.

[38] J. Basque, « Introduction à l'ingénierie pédagogique », Montréal Canada: Université Téluq, 2017, p. 25.

[39] M. Lebrun, « 5 facettes pour construire un dispositif hybride », Blog de Marcel. [En ligne]. Disponible sur: http://lebrunremy.be/WordPress/?p=579

[40] J.-P. Pernin et A. Lejeune, « Dispositifs d'apprentissage instrumentés par les technologies : vers une ingénierie centrée sur les scénarios », présenté à Technologies de l'Information et de la Connaissance dans l'Enseignement Supérieur et de l'Industrie, Université de Technologie de Compiègne, oct. 2004, p. 407. Disponible sur: https://edutice.hal.science/edutice-00000730

[41] A. Jézégou, « Formations ouvertes et autodirection de l'appenant », *Savoirs*, vol. 16, nº 1, p. 97-115, 2008, Disponible sur: https://www.cairn.info/revue-savoirs-2008-1-page-97.htm

[42] K. l. Guftason et R. M. Branch, *What is Instructional Design ?* in In R. A. Reiser, & J. V. Dempsey (Eds.), Trends and Issues in Instructional Design and Technology. New Jersey : Pearson Education, 2007.

[43] G. Paquette, M. Léonard, I. De la Teja, et M.-P. Dessaint, « Méthode d'ingénierie d'un système d'apprentissage MISA 4.0. Présentation de la méthode », 2000.

[44] M. Poumay et F. Geroges, *Comment mettre en œuvre une approche par compétences dans le supérieur ?*, De boeck. in Guides pratiques former et se former. 2022.

[45] D. Ruocco, J.-M. Scheiwiler, et A. Sotnykova, « La qualité des données : concepts de base e techniques d'amlioration », in *La qualité et la gouvernance des données : au service de la performance des entreprises*, Lavoisier, 2012.

[46] J. Brooke, « SUS: A quick and dirty usability scale », *Usability Eval. Ind.*, vol. 189, nov. 1995.

Impact des feedbacks immédiats sur les apprentissages en chimie dans un environnement de réalité virtuelle

Valentin Midez[1], Audrey Serna[1][0000−0003−1468−9761], Florent Diet[1], Victor Oliva Torres[1], and Élise Lavoué[1,2][0000−0002−2659−6231]

[1] INSA Lyon, CNRS, Universite Claude Bernard Lyon 1, LIRIS, UMR5205, 69621 Villeurbanne, France
prenom.nom@insa-lyon.fr
[2] Université Jean Moulin Lyon 3, iaelyon School of Management
elise.lavoue@univ-lyon3.fr

Abstract. La Réalité Virtuelle (RV) permet de créer des environnements immersifs favorisant l'apprentissage pratique en toute sécurité, mais l'immersion des apprenants peut limiter leur réflexion critique. Cette étude explore l'impact de feedbacks immédiats sur l'apprentissage en chimie dans un environnement immersif. Elle implique 46 participants répartis en deux conditions, avec ou sans feedbacks immédiats, dont nous avons observé les performances, incluant les réponses correctes et les procédures réalisées, ainsi que la charge cognitive perçue. Les résultats montrent que les feedbacks immédiats n'augmentent pas la charge perçue, mais que leur effet sur les apprentissages dépend des tâches réalisées. Ils peuvent nuire aux performances pour des tâches ponctuelles, tandis qu'ils conduisent à des améliorations significatives pour des tâches répétées. Ces observations mettent en évidence leur intérêt dans des scénarios répétitifs, tout en soulignant la nécessité d'une conception adaptée pour éviter des effets défavorables sur les apprentissages.

Keywords: Réalité virtuelle, Feedback, Réflexivité, Apprentissage immersif, Enseignement de la chimie.

Abstract. Virtual Reality (VR) offers the opportunity to create immersive environments for safe, practice-based learning, but learners' immersion can limit critical thinking. This study explores the impact of immediate feedback on chemistry learning in a VR environment. It involved 46 participants divided into two conditions, with and without immediate feedback. We assessed their performance (correct answers and procedures performed) and perceived cognitive load. The results show that immediate feedback does not increase the cognitive load, but its effect on learning depends on the task. They can be detrimental to performance for a single task, whereas they lead to significant improvements for repeated tasks. These observations highlight their usefulness in repetitive scenarios while underlining the need for appropriate design to avoid adverse effects on learning.

Keywords: Virtual reality, Feedback, Reflection, Immersive learning, Chemical education.

1 Introduction

La Réalité Virtuelle (RV) permet de simuler des environnements complexes et interactifs, offrant une expérience immersive et engageante propice aux activités d'apprentissage et de formation [3]. Des études montrent que la RV est particulièrement bénéfique pour l'apprentissage actif et l'application pratique [5]. Elle offre un cadre sécurisé où les apprenants peuvent expérimenter et développer des compétences pratiques sans risque, tout en renforçant leur engagement. Les simulations immersives proposent des expériences incarnées réalistes et répétables permettant aux apprenants d'explorer diverses stratégies de gestion de situations complexes et de développer leurs compétences [12].

Cependant, d'autres travaux ont montré que l'immersion créée par ce type d'environnement peut gêner la prise de recul des apprenants [11]. Pour améliorer l'efficacité de ces environnements dans les apprentissages, Roussou et Slater [17] soulignent l'importance de soutenir la réflexivité des apprenants durant la réalisation des activités d'apprentissage. Pour répondre à cette problématique, plusieurs études ont exploré l'intégration de feedbacks immédiats, montrant des résultats positifs sur l'engagement, la réflexion critique des apprenants et leur performances [14,13,1]. D'autres études soulignent cependant un impact plus mitigé des feedbacks sur l'expérience ressentie et notamment sur la charge cognitive perçue par les apprenants selon la complexité des informations fournies par ces feedbacks [21,15]. La conception de ces feedbacks et leur rôle spécifique dans les apprentissages constituent des enjeux cruciaux à approfondir.

Dans cet article, nous nous intéressons à la formation en chimie dont une partie des compétences concerne la pratique d'expériences et de manipulations en laboratoire qui peuvent être travaillées dans des environnements immersifs. Dans ce contexte particulier, nous proposons d'explorer si l'intégration de feedbacks immédiats dans un environnement d'apprentissage en RV a une influence sur l'expérience d'apprentissage. Pour répondre à cette question, nous avons mené une étude comparative auprès de 46 participants répartis en deux conditions expérimentales (environnement avec feedbacks immédiats et sans feedbacks immédiats). Nous avons analysé l'impact des feedbacks à la fois sur les performances (réponses attendues dans le scénario et procédures réalisées) et sur la charge cognitive perçue par les apprenants.

Nos résultats montrent une expérience perçue similaire entre les deux conditions puisque l'ajout de feedbacks immédiats dans l'environnement n'augmente pas la charge cognitive perçue par les apprenants. Concernant les performances, l'effet des feedbacks est différent selon la nature des tâches composant l'activité d'apprentissage. Pour des tâches uniques, les feedbacks immédiats semblent nuire aux performances en termes de réponses correctes attendues seulement. Au contraire, pour des activités mobilisant des tâches répétées, les feedbacks immédiats améliorent les performances avec l'augmentation du nombre de réponses

correctes et l'amélioration de certaines procédures. Ce travail montre que l'intégration des feedbacks immédiats dans les environnements d'apprentissage en RV peut être intéressante selon le scénario d'apprentissage mais qu'il doit se faire avec prudence.

2 État de l'art

2.1 Réalité virtuelle en éducation

En éducation, les dispositifs de réalité virtuelle se distinguent par leurs propriétés immersives et interactives, offrant des environnements dans lesquels les apprenants peuvent pratiquer sans risques physiques ou éthiques [7,1]. Ce type d'environnements peut renforcer l'attention des apprenants en suscitant un sentiment de présence, défini comme l'expérience subjective d'être réellement dans l'environnement virtuel [17,2]. Ils peuvent augmenter la rétention des connaissances et des compétences procédurales, notamment lorsque les méthodes traditionnelles, telles que les démonstrations physiques ou les simulations en classe, sont difficiles à mettre en place en raison des contraintes matérielles, temporelles ou sécuritaires [1], permettant aux apprenants d'expérimenter librement, et de corriger leurs erreurs sans risques réels [9]. La méta-analyse de Cromley et al. [6] met en évidence que les environnements immersifs en RV favorisent non seulement l'acquisition des connaissances, mais également l'engagement des apprenants, ceci pouvant limiter leur capacité à prendre du recul sur leurs actions. Jiang et Ahmadpour [11] soulignent ainsi que l'immersion intense et l'état de flow atteints en RV réduisent la conscience de soi, entravant ainsi la réflexivité sur les actions réalisées.

Comme le soulignent Elaish et al. [8], l'intégration réussie de la RV nécessite ainsi une conception pédagogique visant à aligner l'immersion avec les objectifs d'apprentissage, tout en s'assurant que les environnements virtuels apportent une réelle valeur éducative. Roussou et Slater [17] recommandent notamment de combiner les propriétés interactives des environnements de RV avec le soutien aux processus réflexifs. Selon eux, les environnements interactifs immersifs, s'ils bénéficient de mécanismes de soutien tels que des feedbacks explicites pour informer, confirmer ou corriger les actions des apprenants, peuvent encourager la réflexion critique et prévenir des erreurs potentielles, tout en réduisant le risque de surcharge cognitive. Ces éléments théoriques soulignent l'intérêt d'examiner les mécanismes permettant de soutenir les processus réflexifs en RV.

2.2 Les processus réflexifs et les feedbacks en réalité virtuelle

Les processus réflexifs consistent à réfléchir à la manière dont la tâche a été réalisée, afin de réévaluer les objectifs fixés et les stratégies pour les atteindre [24]. En réfléchissant sur leurs propres apprentissages, les apprenants prennent conscience de leurs processus d'apprentissage et des stratégies alternatives possibles [19]. Ce travail d'introspection permet d'ajuster ses pratiques, d'en développer

de nouvelles ou de redéfinir leur sens [23]. Schön [18] enrichit cette notion en distinguant la réflexion-dans-l'action, qui favorise des ajustements immédiats, et la réflexion-sur-l'action, permettant une analyse critique rétrospective telle que définie précédemment. La réflexion-dans-action peut entraîner la construction de nouvelles compréhensions afin d'ajuster ses actions en temps réel dans la situation qui se déroule.

Une manière de soutenir les processus réflexifs est l'intégration de feedbacks lors de la réalisation de la tâche (réflexion-dans-action) ou après sa réalisation (réflexion-sur-l'action). Hattie et Timperley [16] identifient trois niveaux de feedback : sur la tâche (informations spécifiques sur les performances), sur les processus (orientation sur les stratégies employées), et d'autorégulation (soutien à une évaluation autonome). Dans cet article, nous nous intéressons au premier niveau de feedback lié à la réalisation de tâches d'apprentissage en RV. Plusieurs revues systématiques de la littérature ont été menées tout récemment sur ce sujet. Lee et al. [14] mettent en avant l'importance des feedbacks dans les serious games en RV, augmentée et mixte, pour favoriser la réflexivité, renforcer l'engagement des apprenants et soutenir le développement de compétences cognitives complexes. De même, Lampropoulos et al. [13] montrent que l'utilisation d'environnements d'apprentissage intégrant à la fois la RV et la ludification permet d'améliorer les résultats d'apprentissage et l'efficacité personnelle des apprenants. L'intégration d'éléments comme les niveaux, badges et scores, favorise la réflexion critique en offrant aux apprenants des feedbacks immédiats sur leurs actions.

Quelques études comparatives ont également été menées afin d'étudier l'impact de l'intégration de feedback immédiat sous différentes formes dans des environnements de formation en RV. Almousa et al. [1] montrent que des indicateurs visuels et sonores permettent aux apprenants d'évaluer et d'ajuster immédiatement leurs performances lors de l'apprentissage de gestes techniques en réanimation cardio-pulmonaire. Dans une autre étude sur la formation à l'allaitement, Tang et al. [21] montrent que l'ajout d'une barre de progression, de badges et d'un score, accompagnés de prompts auditifs, permet de maintenir l'engagement global des apprenants tout en guidant leur réflexion sur leurs performances. Enfin, Van Ginkel et al. [22] proposent l'usage de la RV pour le développement de compétences de présentation en public et montrent que des icônes signalant des ajustements nécessaires (comme un contact visuel trop prolongé ou un rythme de parole excessif) facilitent des corrections en temps réel et contribuent à l'amélioration des compétences de présentation.

Cependant, malgré ces apports, des questionnements subsistent quant à l'impact de l'intégration des feedbacks immédiats dans les environnements de RV, notamment sur la charge cognitive induite lors de la réalisation de la tâche, définie comme le rapport entre les demandes spécifiques d'une tâche et les ressources disponibles pour y répondre [20]. Ma et al. [15] montrent que des feedbacks complexes ou multiples peuvent induire une charge cognitive trop élevée qui va limiter les bénéfices pédagogiques. Par ailleurs, Tang et al. [21] montrent une augmentation significative des demandes temporelles et de l'effort requis suscités par la présence de feedbacks immédiats.

En conclusion, si la RV a démontré son efficacité pour favoriser les apprentissages, le rôle spécifique des feedbacks immédiats dans ce type de dispositifs reste encore peu exploré. Par ailleurs, même si les résultats des études menées semblent prometteurs, se pose la question de la charge cognitive induite chez l'apprenant par la visualisation de feedbacks lors de la réalisation de la tâche.

3 Questions de recherche

Dans cette étude, nous cherchons à comprendre si l'ajout de feedbacks immédiats dans un environnement d'apprentissage en réalité virtuelle a une influence sur les apprentissages. Nous nous intéressons plus particulièrement à l'acquisition de compétences en chimie mobilisées dans les manipulations en laboratoire. Nous explorons plus spécifiquement deux questions de recherche :

- QR1 : L'ajout de feedbacks immédiats a-t-il un impact sur les performances des apprenants dans les activités d'apprentissage ? En terme de performances, nous étudions la pertinence des réponses apportées et des procédures réalisées par les apprenants en lien avec les compétences à acquérir via le scénario d'apprentissage.
- QR2 : L'ajout de feedbacks immédiats a-t-il un impact sur la charge cognitive perçue par les apprenants ?

Nous décrivons ci-après l'environnement développé, le scénario d'apprentissage, ainsi que l'intégration des feedbacks pour soutenir la réflexivité.

4 Environnement de réalité virtuelle d'apprentissage de la chimie

Dans le cadre du projet RENFORCE, nous avons développé un environnement d'apprentissage en RV pour la formation en chimie. Une partie des compétences travaillées est orientée sur la pratique des démarches scientifiques : proposer des hypothèses pour répondre à une question scientifique, concevoir des expériences pour les tester, mesurer des grandeurs physiques de manière directe ou indirecte, interpréter des résultats expérimentaux, en tirer des conclusions et les communiquer en argumentant. D'autres compétences portent sur l'adoption de comportements éthiques et responsables, et notamment sur les règles de sécurité et les procédures liées aux manipulations en chimie. Les environnements en RV permettent de travailler ces compétences, notamment les gestes et des manipulations (par exemple, versement d'un liquide d'un bécher à un autre).

L'environnement en RV développé à l'aide de Unity prend la forme d'un laboratoire virtuel, recréant tous les éléments nécessaires pour les manipulations en chimie (plan de travail, sorbonne, robinets, verreries, flacons, matériels de mesure) et intégrant un système de gestion des liquides pour élaborer des mélanges (Figure 1).

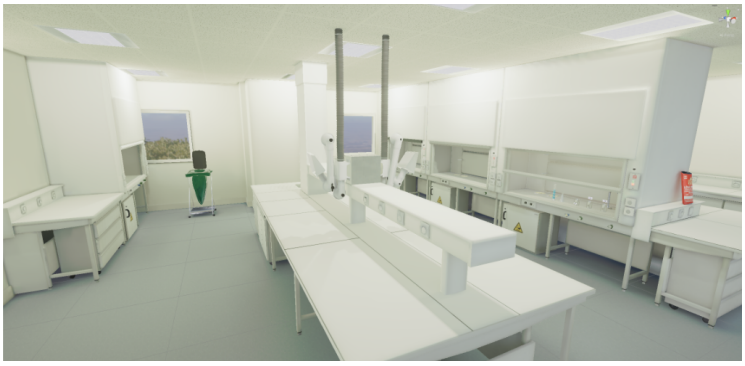


Fig. 1: Environnement virtuel du laboratoire de chimie.

4.1 Scénario d'apprentissage

Pour répondre à nos questions de recherche, nous avons défini un scénario d'apprentissage couvrant des compétences élémentaires de chimie, à savoir observer, expérimenter et mesurer, appliquées au thème de l'organisation et transformation de la matière, plus particulièrement au caractère acide ou basique d'une solution par mesure de pH. Ce thème, travaillé dès le cycle 4, nous permet d'avoir un scénario simple, ne nécessitant pas de prérequis spécifiques, et accessible à différents profils d'apprenants.

La vidéo tutoriel montrée aux apprenants présente les différentes compétences et connaissances couvertes dans le scénario : identifier le caractère acide ou basique selon la présence d'ions H+ et OH- , comprendre les transformations chimiques dans des contextes variés (vie quotidienne, environnement, etc.), utiliser différents moyens de mesure du pH avec du papier pH, avec un pH-mètre ou avec utilisation de colorants, identifier réactions entre solutions acides et basiques et appliquer les procédures appropriées liées aux manipulations.

Le scénario comprend deux manipulations, permettant pour chacune d'évaluer la mobilisation des connaissances et compétences du domaine à travers deux mesures : la pertinence des réponses des apprenants et la réalisation des procédures attendues dans le scénario.

– **Manipulation 1** : les apprenants doivent identifier le caractère acide, basique ou neutre de cinq solutions (eau distillée, café, jus de citron, eau de mer, javel) et utiliser des languettes pH et un tableau de référence pour mesurer et classer chaque solution en fonction de son niveau de pH. Cette étape implique de prélever le liquide avec la pipette, d'appliquer le liquide sur la languette, d'observer le changement de couleur, puis de classer la solution selon sa nature. **Compétences travaillées** : mesurer le pH avec papier pH (languettes), classifier une solution selon le caractère acide, basique ou neutre. **Réponses évaluées** : 2 solutions acides (citron, café), 1 solution neutre (eau distillée), 2 solutions basiques (eau de mer, javel). **Procédure**

à réaliser : utiliser une bandelette par solution, jeter à la poubelle la bandelette pH après utilisation, rincer la pipette après chaque utilisation.

– **Manipulation 2** : les apprenants doivent répéter deux fois une même tâche d'une réaction acido-basique. L'objectif de ces deux tâches est d'atteindre un pH spécifique en utilisant une combinaison de solutions acides et basiques et un indicateur coloré pour évaluer le pH atteint. Dans la première tâche, les apprenants doivent transformer une solution acide (vinaigre blanc) en solution basique de pH 8. Inversement, dans la deuxième tâche, ils doivent transformer une solution basique (ammoniaque) en solution acide de pH 6. Les apprenants doivent choisir la solution titrante (solution basique dans la tâche 1 et solution acide dans la tâche 2) et l'indicateur coloré qui leur permettra de savoir quand le pH cible est atteint. Les apprenants doivent surveiller le changement de couleur du mélange au fur et à mesure du versement de la solution titrante. Ils peuvent utiliser les languettes pH pour vérifier le pH obtenu quand ils le désirent. **Compétences travaillées** : réaliser une transformation en choisissant une solution titrante et un indicateur coloré selon sa zone de virage, atteindre un pH cible. **Réponses évaluées** : indicateur coloré (tâche 1 : rouge de phénol ; tâche 2 : Bleu de bromothymol), solution titrante (tâche 1 : soude ; tâche 2 : acide chlorhydrique), valeur du pH atteint. **Procédure à réaliser** : respecter le bon ordre pour le mélange (indicateur coloré en premier), observer la zone de virage de la couleur du pH, rincer la pipette avant usage, utiliser une bandelette pH, être précis dans l'atteinte du pH, jeter à la poubelle la bandelette pH après utilisation.

4.2 Intégration des feedbacks dans l'environnement

Les feedbacks ont pour objectif d'offrir aux apprenants un retour sur l'acquisition des connaissances et compétences travaillées dans le scénario d'apprentissage et sont construits à partir des réponses évaluées et des procédures spécifiées pour chaque manipulation.

Des feedbacks immédiats combinent des éléments visuels et auditifs pour offrir un retour en temps réel au fur et à mesure des actions des apprenants, afin de soutenir la réflexion-dans-l'action [18] (Figure 2). Pour les concevoir, nous nous sommes inspirés des études récentes proposant des feedbacks immédiats sous forme d'éléments audiovisuels ludifiés [21,13]. En cas de réponse correcte, les feedbacks prennent la forme d'étoiles colorées qui apparaissent accompagnées d'un son symbolisant le scintillement de l'étoile. En cas de réponse incorrecte, les apprenants reçoivent également un retour immédiat sous forme d'étoiles grisées. Concernant les procédures, les feedbacks sont représentés sous la forme d'une barre de progression (ou barre d'XP). Lorsqu'une bonne action au sein d'une procédure est détectée, la barre augmente et s'accompagne d'un son symbolisant cette augmentation.

Pour soutenir la réflexion-sur-l'action [18], nous avons également intégré une synthèse des réponses et procédures réalisées correctement (ou non) à la fin de chaque manipulation (Figure 3). Ce récapitulatif permet aux apprenants de prendre connaissance de leurs performances pour une analyse rétrospective après

chaque manipulation. Cette synthèse est disponible pour l'ensemble des participants quelque soit la condition expérimentale.

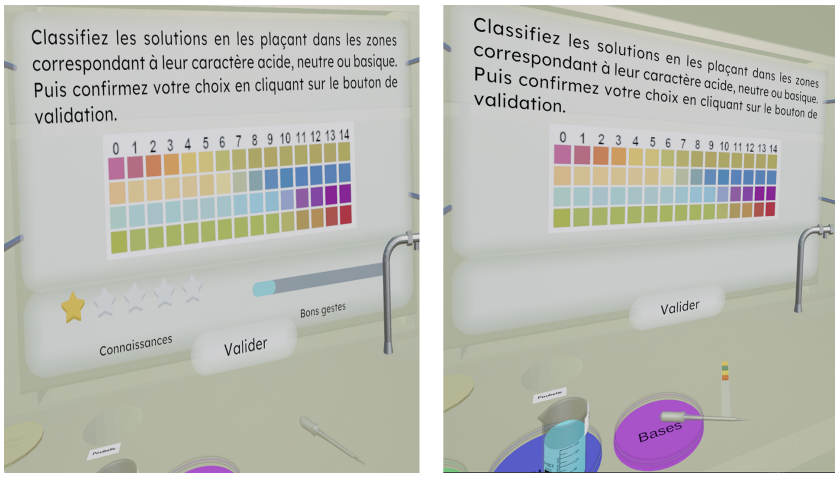


Fig. 2: Panneau de consignes pour chaque condition expérimentale : avec intégration des feedbacks immédiats à gauche (étoiles et barre d'XP) et sans feedbacks immédiats à droite.

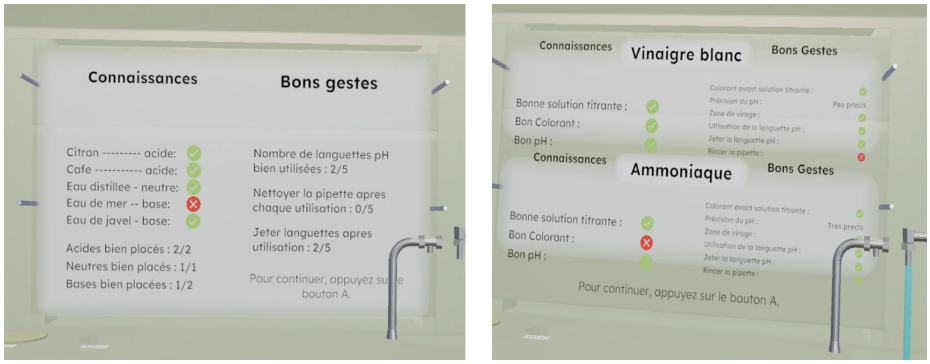


Fig. 3: Synthèse des performances à la fin de la manipulation 1 à gauche et de la manipulation 2 à droite (commune aux deux conditions expérimentales).

5 Protocole de recherche

5.1 Participants

Les participants ont été recrutés via une liste de diffusion universitaire. Les participants ayant une formation ou des études en chimie ont été exclus de l'étude afin d'assurer une homogénéité des connaissances initiales entre les groupes. Un total de 46 participants âgés de 19 à 36 ans (moyenne = 23,4 ans, écart-type = 3,92 ans), ont pris part à cette expérience, dont 17 femmes et 29 hommes. Les participants ont été répartis aléatoirement en deux groupes expérimentaux de 23 participants : un groupe avec feedbacks immédiats (Avec FI) et un groupe sans feedbacks immédiats (Sans FI).

Pour comparer les groupes, des tests non-paramétriques appropriés pour les petits échantillons ont été réalisés (Mann-Whitney pour les valeurs quantitatives et Chi2 pour les valeurs qualitatives). Comme rapporté dans le Tableau 1, les deux groupes sont bien équilibrés selon leurs caractéristiques démographiques, aucune différence significative n'étant observée au niveau de l'âge et du sexe des participants. Les deux groupes ne présentent pas non plus de différences significatives en termes de nombre d'années d'études et d'expertise liée à la RV.

En compensation pour leur temps et leur engagement, chaque participant a reçu une carte cadeau de 10€. Le protocole de recherche a reçu l'approbation du comité d'éthique de la recherche de l'Université de Lyon.

Tableau 1: Répartition des participants selon les variables de genre, âge, années d'études après le bac, fréquence d'expérience en RV, et pratique des jeux vidéos en RV, entre les groupes feedback immédiat et sans feedback immédiat. P-value des tests de Mann-Whitney et Chi-2.

Variables	Catégories	Avec FI (N= 23)	Sans FI (N= 23)	p-valeur
Genre	Homme	14 (61%)	15 (65%)	1.00
	Femme	9 (39%)	8 (35%)	
Âge (Moy ± SD)		23,83 ± 4,93	22,91 ± 2,57	0,80
Années d'études après le bac (Moy ± SD)		4,65 ± 1,67	5,35 ± 1,85	0,17
J'ai déjà porté un casque de RV	Jamais	10 (61%)	8 (47%)	0,64
	Une fois seulement	8 (24%)	11 (33%)	
	Plusieurs fois	4 (12%)	2 (13%)	
	Régulièrement	1 (3%)	2 (7%)	
Je joue à des jeux vidéos en RV	Jamais	18 (62%)	19 (58%)	1.00
	Ponctuellement	5 (38%)	4 (42%)	
	Souvent	0 (0%)	0 (0%)	

5.2 Procédure expérimentale

Une pièce spécialement aménagée, d'une superficie de 6 mètres sur 6 mètres, a été utilisée pour permettre aux participants de se déplacer librement et d'interagir avec l'environnement conçu pour l'étude. Les participants, répartis équitablement entre deux conditions expérimentales, ont suivi une procédure identique. La seule différence entre les groupes consistait en la présence d'un feedback immédiat pendant la réalisation des tâches d'apprentissage.

Avant de commencer, tous les participants ont signé un formulaire de consentement éclairé, conformément aux directives éthiques. Ils ont ensuite rempli un questionnaire démographique, suivi de questions sur leur expérience préalable en RV et leurs habitudes de jeu en réalité virtuelle. Les participants ont également suivi un tutoriel en RV destiné à les familiariser avec le casque de réalité virtuelle (Vive Pro Eye) et les contrôleurs Valve Index. Ce tutoriel comprenait des instructions sur la manipulation d'objets virtuels, la navigation dans l'environnement, et la compréhension du fonctionnement des feedbacks immédiats pour les participants de cette condition. Après ce tutoriel, une vidéo sur les principes acido-basiques en chimie a été montrée aux participants afin de leur fournir les connaissances nécessaires pour réaliser l'expérience en RV. Le visionnage de la vidéo se faisait en dehors de la RV pour éviter d'avoir des temps trop longs en RV. Une fois le tutoriel visualisé, les participants ont de nouveau été équipés du casque de réalité virtuelle pour réaliser les manipulations chimiques dans le laboratoire virtuel. À la fin de la session, les participants ont complété le questionnaire NASA-TLX, afin d'évaluer leur charge cognitive perçue pendant les manipulations réalisées.

5.3 Collecte et analyse des données

Toutes les données d'interaction ont été collectées de façon automatique sous la forme d'actions simples de type « verser une solution dans un bécher » ou « saisir une languette PH ». Ces actions nous ont permis d'analyser les performances des participants durant chacune des manipulations en vérifiant l'exactitude des réponses ou en détectant la réalisation d'une procédure attendue (par exemple « choisir l'indicateur coloré attendu » ou « respecter le bon ordre pour le mélange » pour la manipulation 2). Les performances des participants sont ainsi évaluées à partir du nombre de réponses correctes obtenues dans chacune des manipulations et de la réalisation des actions attendues au sein des procédures à réaliser.

Pour répondre à nos questions de recherche, deux types d'analyses ont été réalisées : une analyse des performances des participants dans les différentes manipulations du scénario (QR1) et une analyse de la charge cognitive perçue (QR2). L'analyse des performances a été réalisée avec Python (*Pandas*, *SciPy*), et l'analyse des réponses au questionnaire avec Jamovi version 2.3. Concernant QR1, les performances des deux groupes ont été comparées en utilisant les scores des réponses correctes et incorrectes, et les scores obtenus à chacune des actions réalisées pour chaque manipulation en utilisant le test de Mann-Whitney. Pour la tâche répétée de la manipulation 2 (tâche 1 et 2), des tests de Wilcoxon,

sans et avec correction de Benjamini-Hochberg, ont été utilisés pour analyser les différences intra-groupes. Étant donnée la taille réduite de nos échantillons (inférieure à 30), nous avons opté pour des tests non paramétriques. Concernant QR2, la charge cognitive perçue a été évaluée à l'aide du questionnaire d'auto-évaluation NASA-TLX [10]. Ce questionnaire évalue six dimensions spécifiques : demande mentale, demande physique, demande temporelle, effort, performance perçue et frustration. Ces dimensions permettent de calculer une charge cognitive globale basée sur les ressentis des participants. Les comparaisons inter-groupes des scores ont été réalisées à l'aide du test non-paramétrique de Mann-Whitney.

6 Résultats

6.1 QR1: Performances sur les réponses attendues

Pour la première manipulation, le groupe avec feedback immédiat obtient un score significativement plus faible de réponses correctes et un score significativement plus élevé de réponses incorrectes comparé au groupe sans feedback immédiat (cf. tableau 2). La taille d'effet modérée confirme que cette différence est notable [4]. Les deux résultats sont rapportés car les participants avaient la possibilité de ne pas répondre.

Tableau 2: Manipulation 1. Moyenne et écart-type des réponses correctes et incorrectes (scores sur 5 points) pour les groupes expérimental (avec FI) et contrôle (sans FI). Résultats du test de Mann-Whitney : p-valeur et taille d'effet (d de Cohen). La significativité à $p < 0{,}05$ est indiquées par *.

	Condition expérimentale	Moyenne (SD)	p-valeur	Taille d'effet
Réponses correctes	Avec FI	4,48 (0,59)	0,03*	0,69
	Sans FI	4,83 (0,39)		
Réponses incorrectes	Avec FI	0,52 (0,59)	0,03*	0,69
	Sans FI	0,17 (0,39)		

Le tableau 3 présente les résultats globaux pour la deuxième manipulation, ainsi que les résultats détaillés de chaque tâche de la manipulation. De façon globale, nous n'observons aucune différence significative entre les deux groupes, qu'ils aient reçu un feedback immédiat ou non. Nous constatons cependant une tendance pour plus de réponses incorrectes pour le groupe avec feedback immédiat durant la tâche 1.

Tableau 3: Manipulation 2. Moyenne et écart-type des réponses correctes et incorrectes (scores sur 6 points, 3 par tâche) pour les groupes expérimental (avec FI) et contrôle (sans FI). Résultats du test de Mann-Whitney : p-valeur et taille d'effet (d de Cohen).

	Réponses	Condition expérimentale	Moyenne (SD)	p-valeur	Taille d'effet
Manipulation entière	Correctes	Avec FI	4,17 (1,37)	0,72	0,13
		Sans FI	4,35 (1,34)		
	Incorrectes	Avec FI	1,61 (1,03)	0,12	0,47
		Sans FI	0,13 (1,01)		
Tâche 1	Correctes	Avec FI	1,78 (0,85)	0,25	0,40
		Sans FI	2,09 (0,67)		
	Incorrectes	Avec FI	1,09 (0,67)	0,08	0,54
		Sans FI	0,74 (0,62)		
Tâche 2	Correctes	Avec FI	2,39 (0,84)	0,91	0,14
		Sans FI	2,26 (1,05)		
	Incorrectes	Avec FI	0,52 (0,67)	0,34	0,19
		Sans FI	0,39 (0,72)		

Cependant, les analyses intra-groupes présentées dans le tableau 4 révèlent des résultats intéressants pour le groupe avec feedback. Les tests post-hoc corrigés selon Benjamini-Hochberg montrent une amélioration significative des réponses correctes entre les deux tâches, ainsi qu'une réduction significative des erreurs. En revanche, aucune évolution significative n'est constatée pour le groupe sans feedback.

Tableau 4: Manipulation 2. Résultats des tests post-hoc de Wilcoxon entre la tâche 1 et 2 : p-valeur brute, p-valeur corrigée (Benjamini-Hochberg) et taille d'effet (d de Cohen). La significativité à $p < 0,05$ est indiquées par *.

	Condition expérimentale	p-valeur brute	p-valeur corrigée	Taille d'effet
Réponses correctes	Avec FI	0,001*	0,022*	0,72
	Sans FI	0,555	0,555	0,20
Réponses incorrectes	Avec FI	0,007*	0,022*	0,85
	Sans FI	0,080	0,100	0,52

6.2 QR1 : Performances sur les procédures réalisées

Nous avons analysé les performances relatives aux procédures pour les deux manipulations. Sans considérer les tâches séparément, nous n'observons pas de différences significatives entre les deux groupes, que ce soit pour la première ou la deuxième manipulation.

Le tableau 5 détaille les résultats pour chacune des deux tâches de la manipulation 2. Avant correction, des différences significatives sont relevées pour les actions *Observer la zone de virage* lors de la tâche 1, en faveur du groupe avec feedback, et *Utiliser une bandelette pH*, lors de la tâche 2, en faveur du groupe avec feedback, avec des tailles d'effet moyennes. Après correction de Benjamini-Hochberg, ces différences ne sont plus significatives, bien qu'elles restent proches de la significativité.

Enfin, les analyses intra-groupes présentées dans le tableau 6 mettent en évidence une différence significative pour l'action *Être précis dans l'atteinte du pH* dans le groupe avec feedback avant correction, montrant une amélioration entre la tâche 1 et la tâche 2, avec une taille d'effet importante. Après correction de Benjamini-Hochberg, cette différence reste quasi significative. Aucune évolution significative n'est constatée pour le groupe sans feedback.

Tableau 5: Manipulation 2. Comparaisons inter-groupes des scores obtenus pour les procédures des tâches 1 et 2 (échelle de 0 à 1). Résultats du test de Mann-Whitney : p-valeurs brutes et corrigées (Benjamini-Hochberg) et tailles d'effet (d de Cohen). Les étoiles (*) indiquent une significativité à $p < 0{,}05$.

Action	Tâche	Avec FI	Sans FI	p-valeur	p-valeur	Taille
		Moyenne (SD)	Moyenne (SD)	brute	corrigée	d'effet
Respecter l'ordre du mélange	T 1	0,86 (0,34)	0,92 (0,22)	0,91	1,00	0,23
	T 2	0,88 (0,29)	0,87 (0,29)	0,55	0,66	0,04
Observer la zone de virage	T 1	0,65 (0,49)	0,30 (0,47)	0,02*	0,12	0,73
	T 2	0,57 (0,51)	0,57 (0,51)	1,00	1,00	0,00
Rincer la pipette	T 1	0,87 (0,34)	0,78 (0,42)	0,45	0,98	0,23
	T 2	0,74 (0,45)	0,57 (0,51)	0,23	0,45	0,36
Utiliser une bandelette pH	T 1	0,91 (0,29)	0,91 (0,29)	1,00	1,00	0,00
	T 2	1,00 (0,00)	0,83 (0,39)	0,04*	0,23	0,63
Être précis dans l'atteinte du pH	T 1	0,17 (0,39)	0,26 (0,45)	0,49	0,98	0,21
	T 2	0,52 (0,51)	0,26 (0,45)	0,08	0,23	0,54
Jeter la bandelette pH usagée	T 1	0,78 (0,42)	0,78 (0,52)	0,95	1,00	0,00
	T 2	0,74 (0,45)	0,65 (0,49)	0,54	1,00	0,19

Tableau 6: Manipulation 2. Comparaisons intra-groupes des scores obtenus pour les procédures des tâches 1 et 2 (échelle de 0 à 1). Résultats du test de Wilcoxon : p-valeurs brutes et corrigées (Benjamini-Hochberg) et tailles d'effet (d de Cohen). Les étoiles (*) indiquent une significativité à $p < 0{,}05$.

Action	Condition expérimentale	Tâche 1 Moyenne (SD)	Tâche 2 Moyenne (SD)	p-valeur brute	p-valeur corrigée	Taille d'effet
Respecter l'ordre du mélange	Avec FI	0,86 (0,34)	0,88 (0,29)	0,93	0,93	0,07
	Sans FI	0,92 (0,22)	0,87 (0,29)	0,47	0,57	0,21
Observer la zone de virage	Avec FI	0,65 (0,49)	0,57 (0,51)	0,56	0,84	0,17
	Sans FI	0,30 (0,47)	0,57 (0,51)	0,08	0,37	0,53
Rincer la pipette	Avec FI	0,87 (0,34)	0,74 (0,45)	0,28	0,55	0,76
	Sans FI	0,78 (0,42)	0,57 (0,51)	0,12	0,37	0,72
Utiliser une bandelette pH	Avec FI	0,91 (0,29)	1,00 (0,00)	0,16	0,49	0,43
	Sans FI	0,91 (0,29)	0,83 (0,39)	0,40	0,57	0,05
Être précis dans l'atteinte du pH	Avec FI	0,17 (0,39)	0,52 (0,51)	0,02*	0,09	0,77
	Sans FI	0,26 (0,45)	0,26 (0,45)	1,00	1,00	0,00
Jeter la bandelette pH usagée	Avec FI	0,78 (0,42)	0,74 (0,45)	0,74	0,89	0,10
	Sans FI	0,78 (0,52)	0,65 (0,49)	0,42	0,57	0,26

6.3 QR2 : charge cognitive

La comparaison des scores du questionnaire NASA-TLX entre les deux conditions expérimentales montre qu'aucune différence significative n'a été observée pour l'ensemble des variables mesurées : pression mentale, pression physique, pression temporelle, effort, frustration, et performance auto-évaluée (cf. tableau 7). Les p-valeurs obtenues restent largement supérieures au seuil de significativité. Les tailles d'effet, faibles pour toutes les dimensions, confirment l'absence d'impact des feedbacks immédiats sur la charge cognitive perçue.

Tableau 7: Comparaison des scores NASA-TLX entre les deux conditions expérimentales. Résultats du test de Mann-Whitney : p-valeur et taille d'effet (r bisérial de rang)

Variable	Condition expérimentale	Moyenne (SD)	Médiane	p-valeur	Taille d'effet
Pression mentale	Sans FI	11,35 (3,81)	12,00	0,82	0,04
	Avec FI	11,09 (4,94)	12,00		
Pression physique	Sans FI	2,91 (3,88)	2,00	0,77	0,05
	Avec FI	2,70 (3,44)	1,00		
Pression temporelle	Sans FI	3,52 (4,07)	3,00	0,27	0,18
	Avec FI	2,48 (3,74)	0,00		
Performance	Sans FI	13,61 (3,66)	14,00	0,80	0,05
	Avec FI	14,09 (3,29)	14,00		
Effort	Sans FI	5,96 (2,75)	6,00	0,38	0,15
	Avec FI	5,43 (4,26)	5,00		
Frustration	Sans FI	14,30 (5,27)	16,00	0,18	0,23
	Avec FI	12,13 (5,86)	13,00		

7 Discussion et conclusion

Dans cette étude, nous avons examiné l'impact des feedbacks immédiats sur les performances des participants dans un environnement immersif en réalité virtuelle (RV) dédié aux manipulations chimiques. Elle a permis d'approfondir les résultats des études précédentes concernant le rôle des feedbacks immédiats sur les performances [1,21,22], en différenciant les effets en termes de réponses et procédures attendues. Les résultats tendent à nuancer les effets positifs obtenus dans ces précédentes études.

Les résultats indiquent que les feedbacks immédiats, tels qu'ils ont été conçus dans cette étude, peuvent être intégrés sans provoquer une surcharge cognitive, même dans des tâches nécessitant une attention particulière et une précision dans les procédures. Ce résultat est en contradiction avec ceux de l'étude menée par Tang et al. [21] montrant une augmentation des demandes temporelles et de l'effort requis. Cela pourrait s'expliquer par l'information apportée par les feedbacks, ceux proposés dans [21] présentant du texte superposé aux éléments eux-mêmes, augmentant ainsi leur complexité [15].

Concernant les performances, les résultats révèlent des effets différenciés des feedbacks immédiats en fonction de la nature des tâches. Pour les tâches ponctuelles, telles que celles de la manipulation 1, la présence de feedbacks immédiats entraîne un nombre de réponses correctes significativement plus faible comparé à celui du groupe sans feedbacks, même si ce résultat n'est pas observé sur les procédures réalisées. Ce résultat suggère que les feedbacks immédiats, en augmentant les informations à traiter simultanément, pourraient réduire les performances dans ce type de tâche. En revanche, pour les tâches répétées de la manipulation 2, les analyses intra-groupes montrent une amélioration significative des réponses correctes et une réduction des erreurs entre les deux tâches pour

le groupe avec feedback immédiat. De même, nous avons observé une amélioration significative (avant correction, quasi significative après correction) pour une action spécifique de la manipulation 2 pour le groupe avec feedbacks immédiats. À l'inverse, aucune amélioration significative n'a été constatée pour le groupe sans feedback. Ces résultats illustrent le rôle des feedbacks immédiats dans l'ajustement progressif des actions et l'acquisition des procédures au fil des répétitions.

En conclusion, bien que la présence de feedbacks immédiats puisse limiter les performances dans des tâches ponctuelles, nos résultats démontrent leur potentiel pour soutenir les apprentissages pour des tâches répétées sans augmenter la charge cognitive. Ces observations soulignent l'importance d'adapter la conception des feedbacks à la nature des tâches et aux besoins des apprenants pour maximiser leur efficacité dans des contextes éducatifs immersifs. Ce résultat rejoint les conclusions de Tang et al. [21], qui recommandent une approche prudente dans l'intégration de feedbacks, ceux-ci pouvant détourner l'attention des aspects essentiels de l'apprentissage. La question de la conception, notamment de la quantité d'information à inclure dans les feedbacks, mériterait d'être explorée plus en profondeur pour maximiser la balance entre charge mentale et performances.

Plusieurs limitations de cette étude méritent d'être soulignées. D'une part, la taille restreinte de l'échantillon réduit la portée des résultats, et le nombre de répétitions des tâches n'est pas suffisant pour explorer pleinement les effets cumulatifs des feedbacks immédiats. D'autre part, le profil des participants et la spécificité du domaine d'apprentissage et des tâches réalisées ne permettent pas une généralisabilité des résultats obtenus à d'autres contextes. Cependant, cette étude ouvre plusieurs perspectives de recherche, notamment l'étude de l'effet de la conception de dispositifs éducatifs en fonction de la nature des tâches, en explorant l'utilisation optimale des feedbacks immédiats pour des tâches répétées. Des recherches ultérieures devraient également se concentrer sur l'impact de la répétition des tâches et sur le rôle des feedbacks dans la rétention des connaissances à long terme, tout en intégrant des scénarios variés pour répondre à des besoins éducatifs diversifiés.

Remerciements. Ce travail a été réalisé dans le cadre du projet ANR-22-FRAN-0005 Ludimoodle+ (financement ANR France 2030). Les auteurs tiennent à remercier tous les participants pour leur participation à l'étude.

References

1. Almousa, O., Prates, J., Yeslam, N., Mac Gregor, D., Zhang, J., Phan, V., Nielsen, M., Smith, R., Qayumi, K.: Virtual Reality Simulation Technology for Cardiopulmonary Resuscitation Training: An Innovative Hybrid System With Haptic Feedback. Simulation & Gaming **50**(1), 6–22 (Feb 2019). https://doi.org/10.1177/1046878118820905
2. Bachmann, M., Subramaniam, A., Born, J., Weibel, D.: Virtual reality public speaking training: effectiveness and user technology acceptance. Frontiers in Virtual Reality **4** (Sep 2023). https://doi.org/10.3389/frvir.2023.1242544

3. Checa, D., Bustillo, A.: A review of immersive virtual reality serious games to enhance learning and training. Multimedia Tools and Applications **79**(9), 5501–5527 (Mar 2020). https://doi.org/10.1007/s11042-019-08348-9
4. Cohen, J.: Statistical power analysis for the behavioral sciences, Rev. ed. Statistical power analysis for the behavioral sciences, Rev. ed., Lawrence Erlbaum Associates, Inc, Hillsdale, NJ, US (1977), pages: xv, 474
5. Conrad, M., Kablitz, D., Schumann, S.: Learning effectiveness of immersive virtual reality in education and training: A systematic review of findings. Computers & Education: X Reality **4**, 100053 (2024). https://doi.org/10.1016/j.cexr.2024.100053
6. Cromley, J.G., Chen, R., Lawrence, L.: Meta-Analysis of STEM Learning Using Virtual Reality: Benefits Across the Board. Journal of Science Education and Technology **32**(3), 355–364 (Jun 2023). https://doi.org/10.1007/s10956-023-10032-5
7. Di Pasquale, V., Cutolo, P., Esposito, C., Franco, B., Iannone, R., Miranda, S.: Virtual Reality for Training in Assembly and Disassembly Tasks: A Systematic Literature Review. Machines **12**(8), 528 (Aug 2024). https://doi.org/10.3390/machines12080528
8. Elaish, M.M., Yadegaridehkordi, E., Ho, Y.S.: Publication performance and trends in virtual reality research in education fields: a bibliometric analysis. Multimedia Tools and Applications (Apr 2024). https://doi.org/10.1007/s11042-024-19238-0
9. Gloy, K., Weyhe, P., Nerenz, E., Kaluschke, M., Uslar, V., Zachmann, G., Weyhe, D.: Immersive Anatomy Atlas: Learning Factual Medical Knowledge in a Virtual Reality Environment. Anatomical Sciences Education **15**(2), 360–368 (2022). https://doi.org/10.1002/ase.2095
10. Hart, S.G., Staveland, L.E.: Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In: Hancock, P.A., Meshkati, N. (eds.) Advances in Psychology, Human Mental Workload, vol. 52, pp. 139–183. North-Holland (Jan 1988). https://doi.org/10.1016/S0166-4115(08)62386-9
11. Jiang, J., Ahmadpour, N.: Beyond Immersion: Designing for Reflection in Virtual Reality. In: Proceedings of the 33rd Australian Conference on Human-Computer Interaction. pp. 208–220. OzCHI '21, Association for Computing Machinery, New York, NY, USA (Sep 2022). https://doi.org/10.1145/3520495.3520501
12. Kiegaldie, D., Shaw, L.: Virtual reality simulation for nursing education: effectiveness and feasibility. BMC Nursing **22**(1), 488 (Dec 2023). https://doi.org/10.1186/s12912-023-01639-5
13. Lampropoulos, G., Kinshuk: Virtual reality and gamification in education: a systematic review. Educational technology research and development **72**(3), 1691–1785 (Jun 2024). https://doi.org/10.1007/s11423-024-10351-3
14. Lee, L.K., Wei, X., Chui, K.T., Cheung, S.K.S., Wang, F.L., Fung, Y.C., Lu, A., Hui, Y.K., Hao, T., U, L.H., Wu, N.I.: A Systematic Review of the Design of Serious Games for Innovative Learning: Augmented Reality, Virtual Reality, or Mixed Reality? Electronics **13**(5), 890 (2024). https://doi.org/10.3390/electronics13050890
15. Ma, X., Wijewickrema, S., Zhou, Y., Zhou, S., O'Leary, S., Bailey, J.: Providing Effective Real-Time Feedback in Simulation-Based Surgical Training. In: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D.L., Duchesne, S. (eds.) Medical Image Computing and Computer-Assisted Intervention MICCAI 2017. pp. 566–574. Springer International Publishing, Cham (2017). https://doi.org/10.1007/978-3-319-66185-8_64

16. Mandouit, L., Hattie, J.: Revisiting "*The Power of Feedback*" from the perspective of the learner. Learning and Instruction **84**, 101718 (Apr 2023). https://doi.org/10.1016/j.learninstruc.2022.101718
17. Roussou, M., Slater, M.: Comparison of the Effect of Interactive versus Passive Virtual Reality Learning Activities in Evoking and Sustaining Conceptual Change. IEEE Transactions on Emerging Topics in Computing **8**(1), 233–244 (Jan 2020). https://doi.org/10.1109/TETC.2017.2737983
18. Schon, D.A., DeSanctis, V.: The Reflective Practitioner: How Professionals Think in Action. The Journal of Continuing Higher Education **34**(3), 29–30 (Jul 1986). https://doi.org/10.1080/07377366.1986.10401080
19. Schunk, D., Zimmerman, B.: Self-regulated Learning: From Teaching to Self-reflective Practice. London: The Guilford Press, New-York, London (1998)
20. Sweller, J.: Cognitive Load During Problem Solving: Effects on Learning. Cognitive Science **12**(2), 257–285 (1988). https://doi.org/10.1207/s15516709cog1202_4
21. Tang, K., Gerling, K., Vanden Abeele, V., Geurts, L., Aufheimer, M.: Playful Reflection: Impact of Gamification on a Virtual Reality Simulation of Breastfeeding. In: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. vol. 783, pp. 1–13. CHI '23, New York, NY, USA (Apr 2023). https://doi.org/10.1145/3544548.3580751
22. Van Ginkel, S., Ruiz, D., Mononen, A., Karaman, C., de Keijzer, A., Sitthiworachart, J.: The impact of computer-mediated immediate feedback on developing oral presentation skills: An exploratory study in virtual reality. Journal of Computer Assisted Learning **36**(3), 412–422 (2020). https://doi.org/10.1111/jcal.12424
23. Verlaine, E.d.: Discovering How to Do Reflexivity and Self-Reflexivity: A Longitudinal Empirical Research Findings. Advances in Social Sciences Research Journal **9**(8), 371–394 (Aug 2022). https://doi.org/10.14738/assrj.98.12931
24. Zimmerman, B.J.: Becoming a Self-Regulated Learner: An Overview. Theory Into Practice **41**(2), 64–70 (May 2002). https://doi.org/10.1207/s15430421tip4102_2

Enjeux juridiques de l'utilisation d'une forge logicielle pour la création et le partage d'EIAH et de ressources pédagogiques en primaire et secondaire

Ezra Nellis[1], Thierry Forest[2], Bertrand Marne[2], Fanny Dessainjean[1] et Iza Marfisi-Schottman[2]

[1] Centre Universitaire Rouennais d'Étude Juridique de l'Université de Rouen Normandie (CUREJ)
ezra.nellis1@univ-rouen.fr, dessainjeanfanny08@gmail.com
[2] Le Mans Université, LIUM, 72085 Le Mans, France
thierry.forest@univ-lemans.fr, bertrand.marne@univ-lemans.fr, iza.marfisi@univ-lemans.fr

Résumé. Permettant le partage et la collaboration, le succès des forges logicielles comme GitLab ou GitHub a dépassé le monde des développeurs. Dans le monde de l'éducation, elles sont poussées par le ministère de l'Éducation nationale à travers sa propre forge. Pourtant, leur utilisation par les enseignants pour partager et collaborer sur des EIAH et des ressources éducatives pose des questions juridiques. Nous présentons les textes de loi sur la propriété intellectuelle appliquée aux enseignants qui utilisent une forge pour partager des ressources et des logiciels. Nous abordons la question des droits moraux, patrimoniaux et des licences logicielles. Nous montrons, au travers de cas pratiques, les types de difficultés posées par l'état actuel des textes. Enfin, en plus d'un arbre décisionnel pour les enseignants, nous proposons deux types de solutions, d'une part fondées sur le renforcement du dépôt des œuvres, d'autre part fondées sur l'utilisation de licences mieux conçues.

Mots-clés : EIAH, Ressources Éducatives Libres, Forge, Droit, Licences.

Abstract. Enabling sharing and collaboration, the success of software forges like GitLab or GitHub has spread beyond the world of developers. In education, they are being pushed by the French Ministry of Education through its own forge. Yet their use by teachers to share and collaborate on Technology-Enhanced Learning systems and educational resources raises legal questions. We present what the law says about the question of intellectual property, applied to teachers using a forge to share resources and software. We address the question of moral and patrimonial rights and software licenses. We use case studies to illustrate the types of difficulties raised by the current state of the law. Finally, we propose two types of solutions, one based on strengthening the deposit of works and the other based on the use of improved licenses.

Keywords: TEL, Open Educational Resources, Forge, Law, Licenses

1 Introduction

Les forges logicielles (ex. : GitLab, GitHub) facilitent le travail des développeurs en proposant, d'une part, un espace d'hébergement doté d'un suivi de modifications des documents et, d'autre part, des fonctionnalités sociales renforçant la cohésion de communautés. Ces fonctionnalités peuvent également bénéficier aux non-informaticiens [1, 2]. Par exemple, une communauté grandissante d'enseignants utilise des forges pour créer et partager des ressources pédagogiques [3, 4, 5]. Le ministère de l'Éducation Nationale français a mis en place une forge officielle, la Forge des Communs Numériques Éducatifs ou **LaForgeEdu** [6], à cet effet. En janvier 2025, elle comptait plus de 4000 comptes enseignants et 2300 projets. Parmi ces projets, on retrouve de nombreux EIAH, sous la forme de code informatique. Mais, les enseignants y partagent également d'autres types de ressources pédagogiques, comme des fiches d'exercices, des livres, des diaporamas de cours et des fichiers multimédias (vidéo, audio). Notons que les fichiers textes sont très similaires aux fichiers de code, puisqu'ils sont composés de lignes de texte dans un format spécifique (ex. : Markdown, HTML) [7].

Sur *LaForgeEdu*, il y a des projets (quasi) individuels (ex. : *ChatMD*, cf. la capture d'écran Fig. 1) et des projets collectifs de grande envergure (ex. : *Ada et Zangemann*), des logiciels (ex. : *AutoBD*, *ChatMD*) et des ressources classiques (*CahierDeVacances*, *Ada et Zangemann*).

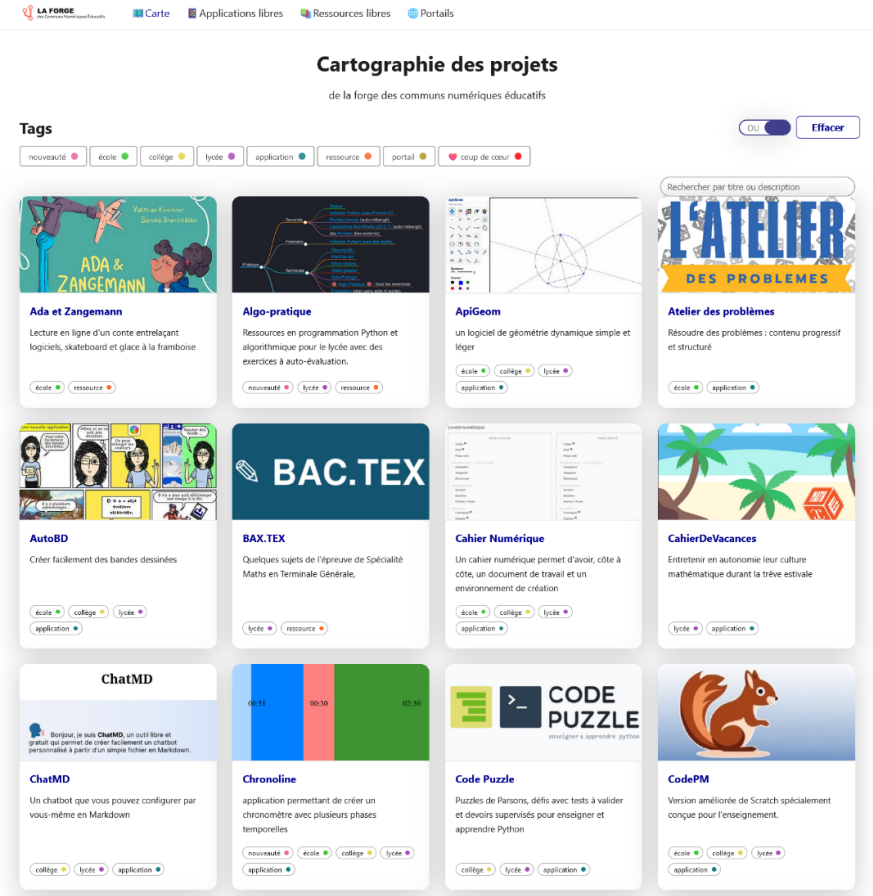


Fig. 1. Quelques exemples de projets sur *LaForgeEdu*

Or, cet usage détourné des forges, comme un environnement de construction, partage et collaboration de ressources pédagogiques, soulève des questions juridiques sur la propriété intellectuelle des ressources créées individuellement ou collectivement par les enseignants du secondaire et du primaire. En effet, juridiquement, la forge, et les contenus qui y sont hébergés sont des **œuvres créatrices de droits**, car ce sont des biens intellectuels ; c'est-à-dire, selon Binctin, des « *choses issues de l'imagination humaine dans l'exercice d'une activité créative susceptible d'appropriation, peu importe son support* » [8]. Il est possible de s'approprier ce type de biens intellectuels selon les règles du régime de la propriété industrielle (ex. : pour les marques, les dessins ou les modèles) ou du régime de la propriété littéraire et artistique (pour les créations intellectuelles qui présentent un certain degré d'originalité, dont font partie les logiciels) [9].

Pour cette raison, la forge logicielle devrait être soumise aux règles inscrites dans le **Code de la Propriété Intellectuelle** (CPI), dont le but est de protéger les intérêts des

individus. Cependant, dans le contexte de contenus pédagogiques sur une forge logicielle, il faut aussi prendre en compte plusieurs autres facteurs. Tout d'abord, comme l'indiquent Blaizot-Hazard et Bernault, le droit français consacre un régime juridique spécifique aux droits patrimoniaux et moraux des agents publics [10, 11], comme les enseignants. De plus, les droits sont différents en fonction du type de ressources pédagogiques. En particulier, les logiciels (ex. : EIAH) ne sont pas soumis aux mêmes règles que des images ou des fichiers d'exercices, par exemple. En outre, le fait que ces ressources soient déposées sur une forge, qui est, elle-même, un logiciel, a des incidences sur le droit qui s'y applique. Enfin, les forges offrent des fonctionnalités pour faciliter la reprise et la modification des ressources d'autres enseignants, mais cette conception collaborative de ressources implique encore d'autres règles juridiques qui ne sont pas forcément compatibles avec les premières. Par conséquent, les règles actuelles qui régissent ses ressources pédagogiques sur les forges ne sont pas claires. Ce flou juridique peut entraver leur partage et donc leur durabilité, mais aussi conduire à leurs détournements (ex. : un usage commercial par une entreprise).

Nos travaux, ancrés ici dans le droit, s'insèrent dans une problématique plus large, courante en EIAH : celle de s'interroger sur la conception, le partage et la collaboration avec des ressources et des logiciels éducatifs dans l'enseignement primaire et secondaire.

Dans cet article, nous présentons un travail exploratoire où nous allons identifier les régimes juridiques actuels qui s'appliquent aux ressources sur les forges logicielles et tenter de démêler leurs points d'incompatibilités. Ce travail est préliminaire à d'autres recherches pouvant conduire à une évolution du droit positif (ex. : propositions de lois).

Il s'agit du fruit d'un **travail interdisciplinaire entre deux juristes et trois informaticiens**. Ces deux disciplines ayant des épistémologies et des pratiques de publication très différentes, nous avons dû faire des aménagements au format habituel des articles en EIAH pour retranscrire au mieux les résultats de nos travaux. En ce sens, dans une première section, nous présentons brièvement notre méthode et sa scientificité. Dans la section 3, nous présentons les lois en vigueur dans le document de référence : le CPI. C'est notre « état de l'art ». Ensuite, dans la section 4, nous présentons notre proposition qui prend la forme d'une interprétation, au plus juste, de ces textes de loi. Nous utilisons plusieurs cas concrets représentatifs des situations de partages d'EIAH ou de ressources que l'on trouve sur *LaForgeEdu*. Pour finir, dans la section 5, apportons des pistes d'améliorations palliant les incertitudes mises en évidence dans les cas concrets.

2 Eléments méthodologiques

La scientificité de cette contribution repose sur l'étude des textes de loi et leur analyse. Cette méthode exégétique est propre aux recherches en droit. Elle mobilise à la fois une terminologie et une démarche spécifiques. Elle repose aussi sur la mise en évidence de cas stéréotypiques problématiques et de leur solution potentielle, avec l'état actuel du droit. Pour la compléter, nous élaborons et proposons des solutions en première approche renforçant la légitimité scientifique de cet article exploratoire. D'autres travaux

suivront avec des solutions plus prospectives (ex. : propositions de loi, amélioration de dispositions existantes, introduction d'éléments plus en adéquation avec les problématiques informatiques).

Nos travaux sont centrés sur le contexte français (notre terrain d'étude) et prennent également en compte leur intégration dans un environnement législatif complexe, européen et international.

3 Ce que dit le droit

3.1 Les droits de propriété intellectuelle

Le CPI est le document de référence qui regroupe les lois régissant **les droits d'auteur**. D'après l'article L.112-2, 13° du CPI (qui ressort du régime d'appropriation de la propriété littéraire artistique), le droit d'auteur s'applique à toute œuvre de l'esprit, comme les ressources pédagogiques. Le créateur de l'œuvre bénéficie de deux catégories de droit :

- **Les droits moraux** garantissent le respect du nom du créateur, de l'œuvre dans son intégrité et de la qualité d'auteur attribué au créateur de celle-ci. Il s'agit d'un droit imprescriptible [12] ! Ainsi, le créateur sera, par exemple, le seul à pouvoir retirer son œuvre du marché, y effectuer des modifications ou s'opposer à des modifications de son œuvre par un tiers. Il sera également le seul à pouvoir accoler son nom à l'œuvre.
- **Les droits patrimoniaux** concernent la sphère économique. Ils donnent, en effet, la possibilité au créateur d'exploiter ou de commercialiser son œuvre. Aussi, il lui est reconnu, par exemple, la possibilité d'autoriser la reproduction de son œuvre ou de la mettre à disposition du public.

Les ressources pédagogiques (ex. : EIAH, livre, fiche d'exercice) qui sont déposées, par les enseignants, sur une forge, sont des œuvres de l'esprit et donnent donc à leurs créateurs des droits d'auteur. Ainsi, il faut se poser la question de savoir à quel type d'œuvre ces contenus pédagogiques appartiennent. Effectivement, le législateur distingue plusieurs types d'œuvres [13] et d'auteurs soumis à des régimes juridiques différents et pas toujours compatibles.

3.2 Le régime juridique spécifique des droits d'auteurs pour les enseignants

Les enseignants du primaire et du secondaire (fonctionnaires ou agents contractuels) ont la particularité d'exercer une mission d'enseignement au service de l'intérêt général et au sein d'une structure éducative. De plus, cette mission d'enseignement participe au bon fonctionnement du service public de l'éducation, comme en atteste notamment l'article L.121-4-, I. du **Code de l'éducation**. Ce statut lie l'enseignant au ministère de l'Éducation Nationale, et, par extension à l'État, soumis aux **règles du droit public**. L'article L. 131-3-1 du CPI estime que « *dans la mesure strictement nécessaire à l'accomplissement d'une mission de service public, le droit d'exploitation d'une œuvre créée par un agent de l'État dans l'exercice de ses fonctions ou d'après les instructions reçues est, dès la création, cédé de plein droit à l'État* ». En d'autres termes, **les droits**

patrimoniaux, spécifiquement le droit d'exploitation, des ressources pédagogiques sont cédées automatiquement à l'État. **Les droits moraux ne sont pas spécifiquement traités** dans l'article susmentionné.

Ce régime s'applique aussi aux enseignants dans les établissements privés, sauf s'ils ont signé un contrat spécifique lors de leur prise de fonction. Seuls les enseignants des établissements publics industriels et commerciaux (ex. : Centres de Formation des Apprentis, CFA) seront soumis aux règles les plus protectrices du droit de la propriété intellectuelle, similaire à celle d'un salarié du privé. Même s'il ne s'agit pas du sujet de cet article, on peut aussi préciser que ce régime ne s'applique pas aux enseignants-chercheurs [14]. Cela s'explique par l'indépendance qui leur est reconnue par l'article L.952-2 du Code de l'éducation.

3.3 Le régime juridique spécifique aux logiciels

Le fait que les ressources pédagogiques soient hébergées sur une forge logicielle implique que ces ressources peuvent être soumises au régime spécifique des logiciels. De plus, ce régime s'applique doublement si la ressource est, elle-même, un logiciel.

En France, selon Binctin, le droit sur les logiciels est très lacunaire et peut être considéré comme peu avantageux pour les développeurs. Il indique que ceci peut s'expliquer par l'absence d'informaticien lors du débat parlementaire pour encadrer les logiciels en 1985 [8]. L'article L.122-6 du CPI reconnaît au créateur de logiciel le droit d'effectuer et d'autoriser la reproduction d'un logiciel, en tout ou partie, par tout moyen et sous toute forme. La reproduction et la modification du logiciel sont également soumises à l'autorisation préalable de son créateur, sauf si ces actes sont commis par des licenciés. C'est-à-dire des établissements scolaires, des académies ou le ministère qui auraient acheté les droits d'exploitation du logiciel. En outre, lorsque le créateur du logiciel est un employé et qu'il agit dans le cadre de ses fonctions, l'article L. 113-9 alinéa 3 du CPI prévoit que les droits patrimoniaux attachés à un logiciel et la documentation sont dévolus à l'employeur. Donc, par principe, **lorsqu'un enseignant crée un EIAH** (ou un autre type de logiciel), **le droit patrimonial sur le code source et la documentation est attribué à l'État**. Le principe reste le même si cette création a lieu en dehors des heures de travail de l'employé s'il utilise le matériel mis à sa disposition par l'employeur [15]. En théorie, cela empêche donc l'enseignant de choisir les modalités de l'exploitation et notamment les licences avec lesquelles il veut déposer son code ou même de percevoir une compensation financière (sauf si un contrat spécifique est signé). Cela dit, dans la pratique, l'État ou ses représentants au niveau académique ne s'occupent pas de déposer ou d'exploiter les EIAH développés par les enseignants. L'état encourage simplement les enseignants à déposer leurs codes sur *LaForgeEdu*.

Quant aux droits moraux, ils ne sont tout simplement pas mentionnés dans ce texte ! Dès lors, qui doit se voir attribuer ces droits : l'enseignant, l'État, le propriétaire de la forge ?

En tout état de cause, les développeurs peuvent faire un dépôt de logiciel auprès de l'Agence pour la Protection des Programmes (APP) [16] pour avoir une preuve de leur paternité, mais ce dépôt coûte au minimum 230 euros et est donc très peu utilisé par les enseignants. Il est aussi possible de recourir aux enveloppes e-soleau de l'INPI [17].

Cette option, moins onéreuse (à partir de 15 euros), donne la possibilité à chacun de déposer des documents dans un espace sécurisé permettant leur archivage ainsi qu'une preuve probante de sa paternité.

Concrètement, un enseignant qui souhaite déposer son code se retrouve donc à faire le choix d'une **licence logicielle** (ex. : Apache, MIT, GNU GPL, BSD, CeCILL. Voir Tableau 1). L'état propose des listes de licences : 15 ciblées par l'article D.323-2-1 du Code des relations entre le public et l'administration [18] et plus de 60 sur un site web officiel de l'État [19]. Cependant, il n'est pas du tout simple pour un enseignant de comprendre les spécificités de ces licences qui, pour la plupart, sont écrites en anglais et, souvent, mal traduites. De plus, une grande majorité de ces licences sont incompatibles avec le droit français. Parmi les sept licences logicielles les plus utilisées (Apache 2.0, MPL, MIT, BSD, EPL, GPL et GPLv3), seule la GPLv3 évoque l'intégrité de l'œuvre. Or, cette mention est indispensable pour respecter le droit français. En effet, selon Clément-Fontaine, la modification du code peut dénaturer la contribution initiale et porter atteinte aux droits du créateur [20]. Dans ce cadre, qu'advient-il du droit de s'opposer à une modification substantielle de son logiciel par le créateur de celle-ci ? En théorie, ce droit lui est pourtant reconnu par le CPI français, puisqu'il est l'auteur de l'œuvre.

Tableau 1. Les droits accordés aux *utilisateurs* pour quelques licences libres typiques par rapport au logiciel propriétaire. Les textes suivis d'une « * » indiquent que le contenu peut être redistribué, *mais seulement sous une licence compatible*.

Licences	Utiliser	Étudier	Modifier	Distribuer tel quel	Distribuer modifié
Propriétaire	Sous condition	Rarement	Rarement	Très rarement	Quasiment jamais
MIT	Oui	Oui	Oui	Oui	Oui
Apache 2.0	Oui	Oui	Oui	Oui	Oui
GPL v3	Oui	Oui	Oui	Oui, mais*	Oui, mais*
CeCILL v2	Oui	Oui	Oui	Oui, mais*	Oui, mais*

3.4 Le régime spécifique aux types d'œuvres

Les forges logicielles permettent aux enseignants d'héberger des ressources pédagogiques de diverses natures (ex. : code, image, fichier texte, fichier audio) qui, en droit, ne sont pas traitées de la même manière. La faculté qu'à la forge logicielle d'héberger des contenus très divers implique de se poser la question de savoir à quel genre d'œuvre appartient-elle ? Les œuvres qui « *réunissent […] sur un support numérique des éléments de genres différents : des sons, des textes, des images fixes ou animées, et des programmes informatiques* » sont des **œuvres dites multimédias** [21]. En raison de la particularité de ces œuvres, les réduire à leur dimension logicielle serait inopportun, car elles se caractérisent par leur interactivité et leur nature numérique. Pour autant, les œuvres multimédias ne bénéficient pas d'une définition juridique dans le CPI. Ce sont

les juges qui ont progressivement permis leur reconnaissance et leur encadrement juridique en les soumettant aux règles du droit d'auteur [22].

De plus, les fonctionnalités offertes par les forges pour produire des ressources de façon collaborative compliquent la définition juridique de ces œuvres. En effet, plusieurs types d'œuvres peuvent convenir. L'article L. 113-2 al. 3 du CPI définit une **œuvre collective** comme « *une œuvre créée sur l'initiative d'une personne physique ou morale qui l'édite, la publie et la divulgue sous sa direction et son nom et dans laquelle la contribution personnelle des divers auteurs participant à son élaboration se fond dans l'ensemble en vue duquel elle est conçue, sans qu'il soit possible d'attribuer à chacun d'eux un droit distinct sur l'ensemble réalisé* ».

Cependant, on peut aussi considérer la forge et ses ressources comme une **œuvre composite**, dans la mesure où l'EIAH ou la ressource qu'un enseignant a déposée sur une forge est une œuvre, à laquelle n'a pas participé le créateur de la forge.

En revanche, si les contributions des enseignants, sur la forge logicielle, sont des ajouts ou des modifications de code qui s'intègrent dans un projet commun et que les utilisateurs ont l'intention de collaborer dans le but de créer un seul produit final, il pourra s'agir d'une **œuvre de collaboration** sur le fondement de l'alinéa 1er de l'article précité. Tel serait le cas, dans un projet open source où des développeurs participent de manière égale à l'ajout de nouvelles fonctionnalités ou à la correction de bugs d'un logiciel via une forge.

Le type de l'œuvre a une incidence sur les droits d'auteur. Par exemple, pour une œuvre de collaboration, l'auteur partage ses droits sur l'œuvre avec les co-auteurs de l'œuvre. De ce fait, ils exercent leurs droits en commun et devront toujours être d'accord entre eux avant de prendre une décision liée à l'œuvre. Par exemple, un seul des co-auteurs ne pourra pas décider seul d'apporter une modification à l'œuvre. Il devra s'assurer au préalable d'obtenir l'accord des autres co-auteurs titulaires. Au contraire, lorsque l'enseignant dépose du contenu sur une forge logicielle, alors la propriété de la personne sous le nom de laquelle l'œuvre est divulguée sera investie des droits d'auteur. De plus, la qualification d'œuvre composite de la forge et de ses données implique que les droits patrimoniaux et moraux des auteurs de l'œuvre première soient respectés quand une seconde œuvre leur est incorporée.

Globalement, comme nous pouvons le constater après cette présentation des lois et des régimes existants, le droit français n'a pas anticipé les questions posées par l'utilisation pédagogique d'une forge logicielle.

4 Quels régimes juridiques pour les EIAH et les ressources pédagogiques sur la forge ?

À travers plusieurs cas concrets, nous allons **proposer une interprétation des lois actuelles** pour clarifier les régimes s'appliquant aux EIAH et aux ressources pédagogiques créées par des enseignants du primaire et secondaire, et partagées sur des forges.

4.1 Cas concret : Partage de fiches d'exercices

Marie, une enseignante de mathématiques, vient de créer des fiches d'exercices sur les fractions et les dépose sur *LaForgeEdu* pour qu'elles puissent être reprises par d'autres collègues. Elle choisit d'y apposer le logo Creative Commons CC By SA (Attribution, Partage dans les mêmes conditions) [23] parce qu'elle a vu d'autres enseignants le faire sur leurs ressources. Elle a créé ces exercices pour sa classe de 6e de collège public, mais sur un ordinateur personnel.

D'un point de vue juridique, ce cas de figure rentre dans le **régime spécifique des agents publics**. En effet, même si le contenu est créé par l'enseignante sur son ordinateur personnel, puis déposé sur la forge de sa propre initiative, les enseignants de primaire et de secondaire font toujours l'objet du contrôle d'inspecteurs dans l'exercice de leur mission auprès de ses élèves. En somme, son autonomie et son indépendance sont toujours relatives. Ce n'est pas l'autonomie, l'indépendance — factuelle ou intellectuelle —, ni l'initiative de l'enseignante qu'il faut retenir comme critère pour savoir si elle peut bénéficier et exercer des droits d'auteur sur son œuvre, mais, c'est son statut (et les règles qui encadrent sa mission d'enseignement) qu'il faut retenir, car elle est toujours soumise à un contrôle dans l'exercice de celle-ci. De plus, les initiatives qu'elle prend pour assumer sa mission d'enseignement sont toujours guidées par les orientations et programmes du ministère. **Les droits de propriété intellectuelle de ses fiches d'exercices sont donc cédés à l'État.** Même si les droits moraux sont donc aussi cédés à l'État, dans la pratique, le nom de la créatrice est mentionné dans le but d'assurer une traçabilité.

Le fait que les ressources soient sur une forge nous oblige à discuter si le régime spécifique aux logiciels s'applique ou non. Si on s'en tient à la définition juridique du logiciel, elle inclut l'ensemble des documents (y compris les ressources) et il faut donc appliquer le régime spécifique des logiciels. Mais, si on considère que les ressources et la forge sont des éléments indépendants, chacune d'elles pourra être soumise à un régime différent prévu par le CPI.

Pour finir, en l'absence de directive, Marie a choisi la licence Creative Commons CC By-SA ce qui semble judicieux, puisqu'il s'agit d'une licence recommandée par l'État. Toutefois, Marie doit bien faire attention à d'autres directives de son établissement qui auraient pu être publiées dans une note de service. La licence choisie stipule clairement que ses ressources peuvent être modifiées et utilisées à condition que son nom soit cité (composante « By » de la licence) et que la nouvelle version soit partagée avec les mêmes droits (composante « SA » de la licence). Par contre, elle ne précise pas que l'on a besoin de son accord pour faire des modifications.

4.2 Cas concret 2 : Modification collaborative de fiches d'exercices

Les fiches exercices de Marie sont très bien faites et une vingtaine d'enseignants de l'association des professeurs de Maths de Lorraine décide de les reprendre. Ils s'inscrivent donc sur *LaForgeEdu* pour créer une nouvelle branche (une dérivation) de ses ressources avec quelques modifications. La plupart des enseignants se sont connectés à la forge via le système d'authentification de l'académie, mais certains ont créé des

comptes externes avec des pseudos du type « Archimed777 ». Ils font bien attention à citer Marie dans la nouvelle version des fiches, et d'y ajouter le logo CC By-SA.

A priori, l'utilisation des fiches d'exercices de Marie par les enseignants de l'association n'est pas guidée par un souhait d'élaborer un projet pédagogique commun avec elle, puisqu'ils ne l'ont même pas contactée. La nouvelle version des fiches ne pourra donc pas être considérée comme une œuvre de collaboration avec Marie. Par contre, il semble que les modifications qu'ils ont opérées ensemble, grâce à *LaForgeEdu*, pourront être qualifiées **d'œuvre composite**. En effet, les fiches d'exercices sont l'œuvre de Marie, à laquelle s'ajoute une seconde œuvre créée par les professeurs de l'association. Ici, l'importance de pouvoir identifier les auteurs des œuvres est fondamentale, car il y a deux œuvres juridiquement indépendantes l'une de l'autre, et cela, même si la seconde s'incorpore à la première. Dans ce cadre, l'utilisation d'un pseudonyme n'est pas recommandée, car il peut priver l'auteur de ses droits. En cas de litige, on peut également se poser la question de savoir si les journaux des traces sur *LaForgeEdu* sont suffisants pour prouver la paternité d'une œuvre. En effet, bien que l'historique d'une forge fondée sur Git ne soit pas modifiable (utilisation des *arbres de Merkle*), ce journal, pour avoir valeur juridique, doit avoir une validation externe (comme avec les *blockchains* ou les *Certificate Transparency* de la RFC 9162) [24-25].

4.3 Cas concret 3 : Partage d'un petit logiciel pédagogique

Camille est une enseignante d'histoire-géo au lycée. Elle a beaucoup d'élèves allophones qui ont du mal à faire les exercices en autonomie. Elle a donc développé un site web qui permet de sélectionner le numéro de ces fiches d'exercices et d'afficher les mots clés et leurs traductions en plusieurs langues. D'un point de vue technique, elle a simplement créé une page web par exercice avec les mots clés. En classe, elle autorise les élèves allophones à utiliser leur smartphone pour accéder à ce site web. En utilisant le navigateur, ils peuvent facilement traduire la page (et les mots clés) dans leur langue natale. Encouragée par ses collègues, elle a mis son code sur *LaForgeEdu* avec une documentation qui explique comment elle l'utilise en classe et comment d'autres enseignants peuvent la modifier. Un peu perdue dans toutes les licences disponibles, elle a choisi la licence MIT parce que c'est une institution qui a une excellente réputation.

Malgré le fait que Marie ait déposé le code seule, sans directive de la part de sa hiérarchie, les droits patrimoniaux de Marie sur son code sont donnés à l'État du fait de son statut d'enseignante. De plus, en souhaitant agir au mieux dans l'intérêt de ses fonctions et de sa mission, Marie prend plusieurs risques en déposant le code avec la licence MIT. En effet, cette licence est permissive et n'assure pas, dans ses termes, une protection complète des droits moraux, notamment l'intégrité de son code qui pourrait être dénaturé par un tiers en le réutilisant. De plus, ce risque s'accroît lorsqu'une licence n'est pas française, car il peut y avoir des incompatibilités avec le droit français. Dans le cas de la licence MIT, l'application de la loi étrangère, moins protectrice que la loi française sur la question du respect de l'intégrité de l'œuvre, complique le respect des droits de Marie garantis par le CPI [26].

4.4 Cas concret 4 : Détournement du logiciel pédagogique à des fins commerciales

Mathieu est un enseignant de technologie. En glanant sur *LaForgeEdu*, il est tombé sur le projet de Camille et a tout de suite eu l'idée de l'exploiter dans son autoentreprise pour les formations professionnelles qu'il donne dans l'entreprise ITEC pendant les vacances. Il a modifié légèrement le code pour ajouter le logo d'ITEC et a modifié le contenu des pages pour afficher « *les informations à retenir* » sur chacun de ces modules de cours.

D'après la licence MIT, Mathieu a tout à fait le droit de reprendre, de modifier et d'exploiter le logiciel de Camille à des fins commerciales. Il n'a même pas besoin d'en informer Camille ! Cependant, Mathieu n'a pas informé son employeur de son autoentreprise, ce qui le met dans l'illégalité. Toutefois, si l'État voulait porter plainte contre Mathieu pour l'usage des ressources, il serait compliqué de prouver l'origine même du code. En effet, Camille n'a pas déposé son code à l'Agence de Protection des Programmes ni à l'INPI (deux services payants), donc il n'y a pas vraiment de preuve qu'elle est l'autrice de l'œuvre. Seul, le dépôt sur *LaForgeEdu* n'est pas une preuve probante au regard de la loi. En effet, même si le dépôt sur cette forge est horodaté et l'auteur est identifié, il est préférable d'avoir une source de registres externe à la forge.

5 Conclusion et perspectives

Comme nous l'avons montré, le droit français est actuellement insuffisant pour répondre clairement aux droits d'auteurs (moraux et patrimoniaux) concernant les ressources pédagogiques déposées sur des forges logicielles. Néanmoins, dans ce travail exploratoire, nous avons pu démêler les différents régimes de droits qui s'appliquent actuellement et proposer une interprétation.

Cependant, comme le montrent les cas concrets présentés dans l'article, il reste de nombreuses zones d'ombre qui apportent de la confusion pouvant amener au détournement de l'œuvre. Il reste encore des problèmes de statut, de traçabilité et d'unification des pratiques pour assurer que les EIAH et les autres ressources pédagogiques puissent évoluer conformément aux souhaits des enseignants.

En parallèle de la mise en place d'outils comme *LaForgeEdu*, il semble donc indispensable de réfléchir aux évolutions juridiques qui vont régir ces usages. En particulier, nous préconisons de **créer un vrai statut pour les forges logicielles**, avec une définition claire, rédigée avec les acteurs des EIAH, qui pourrait être intégrée au **CPI** et dans le **Code de l'éducation**. Le travail pluridisciplinaire entre acteurs des EIAH et juristes, pour faire évoluer le droit, est d'autant plus nécessaire. Cette dynamique participerait à de meilleures transmission et durabilité des innovations pédagogiques.

Nous proposons un diagramme décisionnel pour des utilisateurs des forges et qui peut être mis en œuvre immédiatement (Fig. 2).

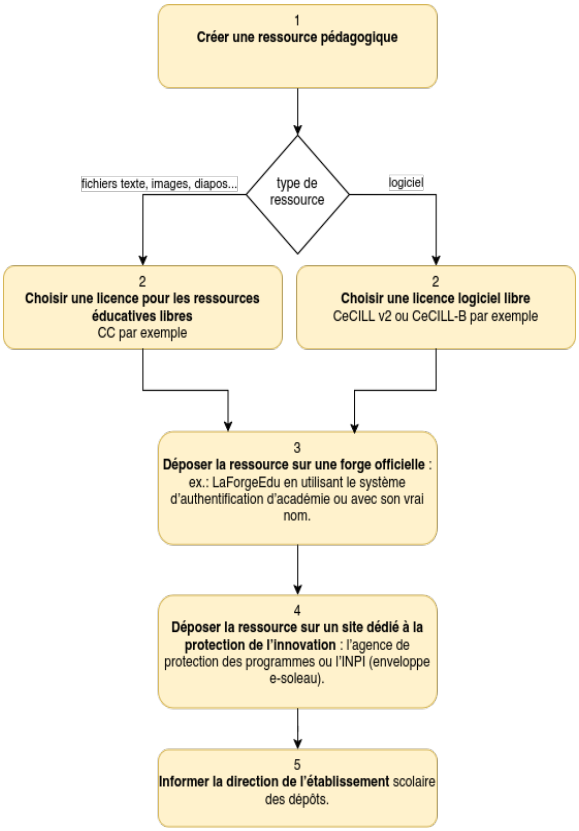


Fig. 2. Arbre décisionnel à destination des enseignants du primaire et du secondaire : comment protéger une ressource mise à disposition sur une forge.

D'autre part, nous proposons deux actions qu'il nous paraît pertinent dans les années à venir : la mise en place d'un système sécurisé pour identifier les auteurs, et une nouvelle licence logicielle, que les enseignants pourraient choisir par défaut.

5.1 Mettre en place une identification sécurisée des co-auteurs

La priorité est de préserver la paternité de l'enseignant sur ses EIAH ou ressources, lorsqu'il recourt à une forge logicielle. Outre d'éviter d'utiliser des pseudonymes, il est proposé de **recourir à un deuxième moyen d'authentification sécurisé du dépôt** de la ressource. La solution de dépôt à l'Agence de Protection des Programmes ou l'INPI (enveloppe *e-soleau*) n'est pas très pratique dans l'usage, car elle est aux frais de l'enseignant, en plus du dépôt sur *LaForgeEdu*. Des « technologies des registres distribués » (ex. : *arbres de Merkle*, socles des forges actuelles) différentes, comme les *blockchains* ou l'association avec des *Certificate Trasparency* [24-25], pourraient être

utilisées pour doubler la sécurisation avec un registre indépendant. Cet usage existe déjà pour répertorier la provenance et la traçabilité des œuvres d'art et éviter leur contrefaçon [27]. Aux yeux de la loi, utiliser une « technologie des registres distribués », indépendante de la forge logicielle, garantit la traçabilité des apports des utilisateurs, et consigne durablement leur identité.

5.2 Proposer une licence logicielle par défaut

La seconde priorité est liée à la nature des licences. C'est un enjeu important, puisqu'en France, la transformation numérique de l'État est activement soutenue par un plan d'action initié par le ministère de la Transformation et de la Fonction publique au cœur duquel se trouvent les logiciels libres [28]. Selon Clément-Fontaine, il existe de nombreuses licences libres pour les codes sources publics et le socle interministériel de logiciels libres ne facilite pas le choix de licence [20]. De plus, aucune de ces licences n'a des conditions harmonisées et toutes ne sont pas compatibles entre elles. Tout cela rend le choix de licence complexe et peu intelligible pour les enseignants. Dans un but de protection des droits des utilisateurs, de simplification et d'intelligibilité, il serait intéressant de se tourner vers un système unifié, voire national. Dans cet esprit, le Québec a instauré une solution innovante et plus lisible [29]. Le gouvernement québécois dispose d'une forge logicielle gouvernementale destinée à « *héberger les projets d'intérêt commun pour les organismes et partenaires gouvernementaux québécois* » et une licence, nommée *LiliQ*, qui a trois versions : LiLiQ-P qui est sans aucune obligation de réciprocité[1] (assimilable à Apache, BSD, MIT), LiLiQ-R qui est une licence avec réciprocité (assimilable à MPL, LGPL) et LiLiQ-R+ qui est une licence à réciprocité forte (assimilable à GPL). LiliQ s'insère harmonieusement dans le paysage du logiciel libre en étant compatible avec la plupart des licences préexistantes [29]. De plus, elle a été conçue en tenant compte du droit de la propriété intellectuelle québécois et donc ne peut pas y déroger.

Références

1. Dodds, T., Reséndez, V., von Nordheim, G., Araujo, T., & Moeller, J.: Collaborative Coding Cultures: How Journalists Use GitHub as a Trading Zone. *Digital Journalism*, *12*(7), 1030–1051 (2024).
2. Dabbish, L., Herbsleb, J.: Open collaborative writing: Investigation of the fork-and-pull model. In: Proceedings of the ACM on Human-Computer Interaction, Volume 5, Issue CSCW1, Article No 137, pp. 1–33 (2010).
3. Glassey, R.: Adopting git/github within teaching: A survey of tool support. In: Proceedings of the ACM Conference on Global Computing Education, pp. 143–149 (2019).
4. Zagalsky, A., Feliciano, J., Storey, M.A., Zhao, Y.: The emergence of github as a collaborative platform for education. In : Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing, pp. 1906–1917 (2015).

[1] Réciprocité veut dire qu'il est impossible pour un licencié de modifier et de distribuer un logiciel sans en dévoiler le code source ou de conserver le caractère libre du logiciel.

5. Santos-Hermosa, G.: The role of institutional repositories in higher education: Purpose and level of openness. Distributed Learning Ecosystems, pp. 47–70 (2023).
6. Page d'accueil de LaForgeEdu, https://docs.forge.apps.education.fr/, consulté le 10/01/2025.
7. Bani-Salameh, H., Jeffery, C.: Collaborative and social development environments: a literature review. International Journal of Computer Applications in Technology, 49(2), pp pp. 89–103 (2014).
8. Binctin, N. : Droit de la propriété intellectuelle, Droit d'auteur, brevet, droits voisins, marque, dessins et modèles, éd. LGDJ, Paris (2024).
9. Article L112-2 du Code de la Propriété Intellectuelle.
10. Blaizot-Hazard, C. : Les droits de propriété intellectuelle des personnes publiques en droit français, bilb. de droit public, t. 159, LGDJ (1991).
11. Bernault, C. : Droit d'auteur - Le droit d'auteur des enseignants - L'enseignant est-il un fonctionnaire « comme les autres » ?, Communication Commerce électronique n° 3, étude 6 (2010).
12. Article L.121-1 al. 3 du Code de la Propriété Intellectuelle.
13. Bensamoun, A., Labarthe, F. Tricoire, A. : L'œuvre de l'esprit en question(s) : un exercice de qualification, éd. Mare et martin, Paris (2015).
14. Terrier, E., Fasc. 1187 : Objet du droit d'auteur. Agents publics (CPI, art. L. 111-1, al. 3), JurisClasseur, Propriété littéraire et artistique, mise à jour, 15 sept. 2024, spé., § 49.
15. Cour d'Appel de Nancy, 13 sept. 1994, JCP E 1996, I.559, n°2 (1994)
16. Page d'accueil de l'Agence pour la Protection des Programmes, https://www.app.asso.fr/, consulté le 10/01/2025.
17. Page web sur le enveloppes e-soleau de l'INPI, https://www.inpi.fr/services-et-prestations/e-soleau, consulté le 10/01/2025.
18. Article D. 323-2-1 du Code des relations entre le public et l'administration, accessible en ligne https://www.data.gouv.fr/fr/pages/legal/licences/
19. Page de la DINUM, https://code.gouv.fr/fr/doc/licences-libres-dinum/, consulté le 10/01/2025.
20. Clément-Fontaine, C. : L'œuvre libre, éd. Larcier, Bruxelles (2014).
21. Montel, B., Fasc. 1140 : Object dur droit d'auteur – Œuvres protégées. Œuvre audiovisuelle (CPI, art. L. 112-2), JurisClasseur Propriété littéraire et artistique, 30 juin (2023).
22. Cour d'appel de Paris, 28 avr. 2000, n° 1999/02618 : JurisData n° 2000-118091 ; Communication. commentaire. 86, observation. Ch. Caron ; RIDA 1 /2001, p. 314 ; D. 2001, p. 2553, observation. P. Sirinelli ; Légipresse 2000, n° 173, III, p. 107, note A. Latreille (2000).
23. Page web de la licence Creative Commons, https://creativecommons.org/licenses/by-sa/4.0/deed.fr, consulté le 10/01/2025.
24. Grilli, L., Speziali, P.: Combining Git and Blockchain for Trusted Information Sharing. IEEE Access. 12, 88383–88409 (2024). https://doi.org/10.1109/ACCESS.2024.3418749.
25. Laurie, B., Messeri, E., Stradling, R.: RFC 9162: Certificate Transparency Version 2.0, (2021).
26. Affaire Huston, Cour de Cassation, 1ère chambre civile, 28 mai 1991, RIDA, juill. 1991, n° 149, p. 197, et Cour d'Appel de Versailles, chambres civiles réunies, 19 décembre 1994, RIDA, avril 1995, p. 369 (renvoi de cassation).
27. Joly, L. : La blockchain est-elle une révolution pour la propriété intellectuelle ?, IP/IT : droit de la propriété intellectuelle et du numérique, 10, 536–539 (2018).
28. Page web de la Mission Logiciel Libre de l'État, https://code.gouv.fr/, consulté le 10/01/2025.

29. Cadre de référence des logiciels libres, Ministère de la cybersécurité et du numérique, Gouvernement de Québec, Bibliothèque et Archives nationales du Québec (2023).

Recommander pour soutenir l'exploration urbaine et le partage d'expériences entre étudiants internationaux

Madjid Sadallah et Marie Lefevre

Universite Claude Bernard Lyon 1, CNRS, INSA Lyon,
LIRIS, UMR5205, 69622, Villeurbanne, France
{madjid.sadallah, marie.lefevre}@liris.cnrs.fr

Résumé. L'intégration des étudiants internationaux, cruciale pour leur inclusion sociale et leur apprentissage, rencontre souvent divers obstacles. Bien que les outils numériques, telles les applications de découverte et de documentation, puissent soutenir ce processus, leur efficacité repose sur une appropriation réfléchie. Cet article présente un système de recommandations personnalisées conçu pour guider les étudiants internationaux dans l'utilisation d'une application d'exploration urbaine et de partage d'expériences. En offrant des suggestions contextuelles, ce système encourage une exploration active et une meilleure compréhension de l'environnement local, renforçant leur intégration sociale et culturelle. Une évaluation empirique a révélé une amélioration de la qualité narrative, de la compréhension du contexte local et de l'expression émotionnelle des étudiants. Ces résultats soulignent le potentiel des systèmes de recommandations pour enrichir l'apprentissage contextuel et faciliter l'intégration, facteurs clés du bien-être et de la réussite académique des étudiants internationaux.

Mots-clé : Systèmes de recommandations · Étudiants internationaux · Exploration urbaine · Partage d'expériences · Annotation spatiale.

Abstract. The integration of international students, crucial for their social inclusion and learning, often faces various obstacles. While digital tools, such as discovery and documentation applications, can support this process, their effectiveness depends on thoughtful adoption. This article introduces a personalized recommendation system designed to guide international students in using an urban exploration and experience-sharing application. By offering contextual suggestions, this system encourages active exploration and a deeper understanding of the local environment, thereby enhancing their social and cultural integration. An empirical evaluation revealed improvements in students' narrative quality, local context comprehension, and emotional expression. These findings highlight the potential of recommendation systems to enrich contextual learning and facilitate integration, key factors for the well-being and academic success of international students.

Keywords: Recommender Systems · International Students · Urban Exploration · Experience Sharing · Spatial Annotation

1 Introduction

L'intégration réussie des étudiants internationaux dans leur nouvel environnement académique et urbain est cruciale pour leur épanouissement et leur réussite [13]. Cette transition soulève cependant des défis majeurs comme les barrières linguistiques, les différences culturelles et l'isolement social [19, 4]. Amplifiés par les disparités entre systèmes éducatifs et les difficultés de communication, ces obstacles peuvent nuire à l'engagement, à la satisfaction, voire à la santé mentale des étudiants. Il est donc essentiel de développer des solutions innovantes pour faciliter leur intégration sociale et améliorer leur bien-être, facteurs clés de leur succès académique.

Les approches traditionnelles, telles que les programmes d'orientation ou le mentorat, bien qu'utiles à court terme [10], montrent des limites pour une intégration durable. La recherche récente souligne le potentiel des outils numériques, notamment les applications mobiles et plateformes en ligne, pour un accompagnement en temps réel [7, 18, 1, 9]. Ces technologies peuvent faciliter la découverte des ressources locales, renforcer les liens sociaux et offrir un soutien personnalisé. Toutefois, ces solutions restent souvent fragmentaires, ciblant des aspects spécifiques au détriment d'une approche holistique des dimensions académiques, sociales et culturelles [19, 1]. Un besoin se fait sentir pour des outils numériques intégrés, flexibles et adaptatifs, afin d'accompagner efficacement les étudiants internationaux [12].

Les systèmes de recommandations personnalisées représentent une voie prometteuse pour répondre à ce défi. Analysant les préférences et comportements des utilisateurs, ils ont prouvé leur efficacité dans divers domaines, y compris l'éducation et l'intégration culturelle [17, 15]. Grâce au filtrage collaboratif ou à la recommandation par contenu, ils proposent des suggestions contextuelles en temps réel, tenant compte des spécificités individuelles. Appliqués à l'intégration des étudiants internationaux, ces systèmes pourraient faciliter la découverte urbaine, l'accès aux ressources et encourager la réflexivité par la documentation des expériences sociales et culturelles [3, 18]. Néanmoins, bien que leur potentiel en éducation soit reconnu, leur application spécifique à l'intégration des étudiants internationaux demeure encore largement sous-explorée.

Dans le cadre du projet interdisciplinaire *MOBILES* [1], nous avons développé une application éponyme permettant aux étudiants internationaux de documenter et partager leurs expériences urbaines et sociales. Les premiers retours ont révélé des difficultés à produire des annotations détaillées, dues notamment aux obstacles linguistiques et culturels. Pour y pallier, nous avons intégré un système de recommandations personnalisées visant un double objectif synergique : (1) stimuler la découverte de l'environnement par des suggestions de lieux pertinents, favorisant l'intégration sociale et culturelle ; et (2) améliorer la qualité et la quantité des annotations, encourageant une documentation plus riche et fréquente. Ces aspects se renforcent mutuellement : la découverte incite à documenter, et des annotations de qualité enrichissent l'exploration. Le

[1] https://mobiles-projet.huma-num.fr/

système propose des suggestions adaptées aux pratiques individuelles pour faciliter l'immersion et enrichir l'expérience.

L'étude présentée dans cet article vise à évaluer l'efficacité de ce système et son impact sur les pratiques documentaires des étudiants internationaux. Bien que le système adresse les deux objectifs susmentionnés, nous nous concentrons ici sur l'évaluation de l'impact des recommandations visant à améliorer la qualité et la quantité des annotations. La question de recherche centrale est : *Comment un système de recommandations personnalisées influence-t-il l'engagement, la diversité des pratiques documentaires et la qualité des annotations produites par les étudiants internationaux lors de leurs expériences urbaines et sociales ?* Pour y répondre, nous avons mené une étude empirique pour analyser l'impact du système de recommandations sur ces trois aspects.

La suite de cet article présente d'abord l'application *MOBILES*, puis décrit le système de recommandations intégré. Nous exposons ensuite la méthodologie d'évaluation, suivie des résultats de notre étude empirique. Enfin, une discussion des résultats et des conclusions clôture l'article.

2 L'application *MOBILES*

MOBILES est une application pour smartphones destinée aux étudiants internationaux afin de documenter et partager leurs expériences sociales et urbaines *via* des annotations détaillées [8]. Elle intègre une carte interactive de la ville visualisant les annotations de lieux et les trajets partagés, offrant une cartographie vivante des expériences urbaines.



(a) Page d'accueil. (b) Une annotation. (c) Un parcours.

Fig. 1: Interface principale de *MOBILES*.

L'interface principale (Figure 1) donne accès à des fonctionnalités clés comme la découverte de lieux, la carte interactive et les annotations des étudiants. L'application comprend des modules pratiques tels qu'un gestionnaire de favoris (pour sauvegarder points d'intérêt, annotations et parcours), le système de recommandations personnalisées, proposant des suggestions pour enrichir l'exploration et un système de notifications informant les étudiants des mises à jour relatives à leurs contenus.

Le processus de création d'annotations (Figure 2) permet aux étudiants de documenter leurs découvertes en précisant des informations temporelles (fréquence, date) et la portée géographique (lieu spécifique, quartier, ville). Chaque annotation peut inclure, au choix, un champ textuel descriptif, des photos, et des icônes catégorisées par couleur couvrant : (1) *Activité* pour les actions principales réalisées, (2) *Lieu* pour les caractéristiques physiques, (3) *Sensoriel* pour les perceptions sensorielles, (4) *Social* pour les interactions humaines, et (5) *Affectif* pour les émotions ressenties. Des tags et émoticônes (Figures 2c et 2d) permettent d'exprimer des ressentis (sécurité, utilité, émotions).



(a) Contexte (b) Liste d'icônes (c) Liste de tags (d) Émoticônes

Fig. 2: Composants essentiels du processus de création d'une annotation.

Les annotations peuvent être publiques, privées ou partagées en groupes, et sont modifiables pour refléter de nouvelles perspectives. Elles peuvent également être associées à des parcours urbains, offrant ainsi un suivi dynamique des expériences. Les étudiants peuvent interagir avec les annotations d'autres utilisateurs en réagissant ou en commentant.

Les parcours (Figure 1c) offrent une vue d'ensemble des explorations en enregistrant trajets et points d'intérêt visités. Ils se créent de trois manières : (1) *Suivi GPS automatique* (enregistrement en temps réel du trajet) ; (2) *Cartographie manuelle* (dessin d'itinéraires) ; et (3) *Séquençage rétrospectif* (liaison d'annotations en un récit narratif). Ces méthodes permettent aux étudiants de construire un récit visuel et dynamique de leurs explorations.

3 Système de Recommandations de *MOBILES*

Afin de faciliter la découverte urbaine et d'enrichir l'expérience des étudiants internationaux, nous avons développé un moteur de recommandations personnalisées pour *MOBILES*. Il analyse les traces des actions des utilisateurs (création d'annotations, consultation de lieux, exploration cartographique, etc.), collectées et gérées par le module *Kernel for Trace-Based Systems* (kTBS) [16], un service web RESTful permettant de gérer des séquences d'observations horodatées. Ce module enregistre les traces des comportements, offrant une vue détaillée des pratiques et préférences. En exploitant ces traces, le moteur de recommandations génère des suggestions personnalisées pour stimuler la découverte et améliorer la documentation. Les détails techniques de l'implémentation sont disponibles dans [14].

3.1 Tâches de recommandations

Pour atteindre ses objectifs, le système met en œuvre deux tâches de recommandations distinctes.

Recommandation de lieux à visiter Cette tâche vise à aider les étudiants à découvrir des lieux pertinents dans leur nouvel environnement, favorisant ainsi leur exploration et leur intégration. Les recommandations de lieux s'appuient principalement sur le *filtrage basé sur le contenu* des annotations et sur la *popularité* des lieux visités.

Recommandation d'incitations à l'annotation. Constituant le focus principal de l'évaluation présentée dans cet article, cette tâche vise à améliorer la qualité et la quantité des annotations, favorisant réflexivité et engagement dans la documentation. Elle repose sur un ensemble d'heuristiques basées sur l'analyse du contenu des annotations existantes de l'utilisateur. Ces heuristiques s'appuient sur les cinq métriques suivantes pour identifier les points d'amélioration potentiels.

Stimuler l'engagement actif. L'engagement dans la documentation est essentiel pour une couverture complète des vécus et un contenu riche pour la réflexion. La métrique *Engagement dans la Documentation (DocEng)* mesure l'activité globale de documentation d'un utilisateur sur une période donnée. Elle est définie comme le nombre total d'actions de création, modification ou édition :

$$\text{DocEng} = C + M + T \tag{1}$$

où C représente le nombre total d'annotations créées, M le nombre total d'annotations modifiées, et T le nombre total de parcours édités par l'utilisateur sur la période considérée. Lorsque *DocEng* est inférieur à un seuil déterminé (actuellement fixe, une approche dynamique est envisagée comme perspective), le moteur de recommandations incite à accroître l'activité de documentation.

Enrichir les récits textuels. L'expression textuelle permet de documenter les expériences tout en soutenant la réflexion et l'analyse critique des situations vécues. Des descriptions riches et significatives renforcent l'engagement et la pensée critique, notamment lorsqu'elles sont encouragées par des incitations personnalisées [11]. La métrique *Richesse Textuelle (TextRich)* évalue cette richesse comme suit :

$$\text{TextRich} = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{LV_i + LD_i}{2} \right) \tag{2}$$

où LV_i représente le volume lexical (nombre total de mots) et LD_i la diversité lexicale (nombre de mots distincts). Si la valeur de *TextRich* d'un étudiant est inférieure à la moyenne de l'ensemble des étudiants, le moteur propose des incitations pour encourager des annotations plus détaillées.

Bien que TF-IDF permette d'évaluer la spécificité des termes dans un corpus, nous avons choisi *TextRich* pour sa simplicité et sa capacité à mesurer la richesse textuelle sans corpus externe. TF-IDF reste néanmoins une piste prometteuse pour analyser les thématiques dans de futurs travaux.

Encourager l'expression émotionnelle. L'intégration d'émotions et de pensées dans les annotations apporte une dimension authentique et personnelle, essentielle pour comprendre le paysage émotionnel des expériences urbaines et leurs effets sur les interactions sociales [2]. La métrique *Indice d'Expression Émotionnelle (EmoExp)* permet d'évaluer cet aspect :

$$\text{EmoExp} = w \cdot \text{AAR} + (1 - w) \cdot \text{AIR} \tag{3}$$

où AAR est le *Taux d'Annotations Affectives* (proportion d'annotations contenant des expressions émotionnelles), AIR est le *Taux d'Icônes Affectives* (ratio des icônes émotionnelles utilisées), et w est un coefficient de pondération attribué à AAR. Dans notre première implémentation, nous avons fixé $w = 0{,}7$, donnant ainsi plus de poids à l'expression textuelle (AAR), jugée plus nuancée que les icônes (AIR). Cette valeur pourrait être affinée par une analyse de sensibilité ou adaptée dynamiquement selon le contexte ou l'utilisateur. Si *EmoExp* passe sous un seuil, le système encourage l'étudiant à enrichir sa contribution sur le plan émotionnel.

Intensifier l'usage graphique. Les éléments graphiques (images, photos) améliorent la compréhension, l'accessibilité et la mémorisation, tout en rendant les annotations plus claires, engageantes et adaptées à des publics variés [5, 6]. La métrique *Richesse Graphique (GraphRich)* évalue cet usage :

$$\text{GraphRich} = \frac{N_{\text{graphics}}}{N_{\text{total}}} \tag{4}$$

où N_{graphics} représente le nombre d'annotations ayant des éléments graphiques, et N_{total} le nombre total d'annotations créées par ce même utilisateur. Si la valeur de GraphRich est inférieure à un seuil défini, le système incite à inclure davantage d'éléments graphiques.

Diversifier les éléments iconographiques. Une variété d'icônes enrichit la narration visuelle, améliore l'accessibilité et aide à surmonter les barrières linguistiques, rendant la documentation plus attrayante. La métrique *Diversité Iconographique (IconDiv)* quantifie cet aspect :

$$\text{IconDiv} = \alpha \cdot ANI + \beta \cdot ANIT \tag{5}$$

où ANI désigne le nombre moyen d'icônes par annotation et $ANIT$ le nombre moyen de types d'icônes différents par annotation pour ce même utilisateur. Les coefficients α et β ajustent l'importance relative de la quantité et de la diversité des icônes, permettant de personnaliser la métrique en fonction des priorités d'analyse. Leur valeur peut être modifiée en fonction des objectifs ou des données empiriques. Lorsque *IconDiv* est inférieur à un seuil, des recommandations sont générées pour accroître la variété des icônes.

3.2 Architecture

Le système repose sur un module de traçage qui enregistre en temps réel les activités des étudiants, intégrant des données géospatiales et horodatées. L'analyse de ces données initie le processus de recommandation, suivi d'un prétraitement comprenant nettoyage, déduplication et standardisation. Des indicateurs comportementaux sont ensuite calculés pour générer des recommandations contextuelles, correspondant aux deux tâches principales (lieux et incitations).

Le moteur de recommandations utilise diverses techniques pour la recommandation de lieux (filtrage de contenu, popularité, analyse géographique) et des heuristiques basées sur les métriques pour générer des incitations à l'annotation. Chaque recommandation inclut généralement un *Prompt* (directive personnalisée, par exemple "*Ajoute plus de détails*") et parfois une *Suggestion* (par exemple un lien vers une annotation pertinente). Les recommandations de lieux sont présentées sous forme de liste de points d'intérêt.

Les recommandations sont transmises par notifications *push* et *via* le centre de notifications de l'application, avec des ajustements de fréquence et de moment d'envoi pour maximiser leur pertinence sans être intrusives. Une boucle de rétroaction recueille les interactions des étudiants, permettant ainsi une amélioration continue du moteur de recommandation.

4 Étude

4.1 Méthodologie et participants

Cette étude visait à évaluer l'impact des recommandations sur les pratiques des étudiants internationaux utilisant *MOBILES* pour explorer la ville de Lyon. L'objectif principal était d'analyser l'influence de ces recommandations sur la qualité des annotations et l'engagement des participants dans leurs expériences urbaines, contribuant à leur intégration sociale. Une approche méthodologique

mixte a été adoptée pour mesurer l'efficacité du système de recommandations et recueillir les retours d'expérience.

Les participants ont été recrutés parmi les étudiants internationaux des différentes universités de Lyon *via* affichages sur les campus et listes de diffusion. L'échantillon final comprenait 31 étudiants (15 hommes, 16 femmes), d'âge médian 24 ans (étendue : 19-43 ans). Neuf étaient en première année d'études, 22 en cours de parcours (14 en licence, 8 en master). Concernant l'origine géographique, 18 participants provenaient d'Afrique, 4 des Amériques, 4 d'Asie et 2 d'Europe.

L'étude s'est déroulée du 25 mars au 2 mai 2024 en trois phases. La première phase a consisté en l'obtention du consentement éclairé, la présentation du projet et l'installation de l'application. Durant la deuxième phase, les participants ont exploré Lyon de manière autonome, documentant leurs expériences avec l'application. Des recommandations personnalisées leur étaient envoyées régulièrement *via* des notifications intégrées pour encourager l'utilisation des fonctionnalités. Les comportements des participants ont été suivis avec leur consentement, permettant de collecter des données détaillées en contexte réel. La troisième phase a permis de recueillir les retours des participants par un questionnaire en ligne et des *focus groups* en présentiel.

4.2 Instruments

Pour évaluer l'impact du système de recommandations sur l'amélioration des annotations, nous avons combiné l'analyse des traces d'utilisation de l'application et la collecte de retours qualitatifs.

L'analyse des traces a porté sur l'évolution des cinq métriques clés sous-tendant les heuristiques de la tâche de recommandation d'incitations à l'annotation (Section 3.1). Cette analyse visait à évaluer l'influence de ces incitations spécifiques sur les pratiques documentaires. Parallèlement, des *focus groups* ont été organisés pour discuter des évolutions observées dans ces métriques, offrant une perspective qualitative complémentaire.

Les perceptions et expériences des participants ont été recueillies en fin d'étude *via* un questionnaire en ligne structuré, divisé en deux sections. La première comportait des questions fermées, évaluées sur une échelle de Likert à cinq points (1 = fortement en désaccord, 5 = fortement d'accord), portant sur :

- *Pertinence :* les recommandations étaient alignées avec les intérêts des participants.
- *Variété :* les recommandations abordaient plusieurs aspects différents et suggéraient une diversité de lieux.
- *Facilitation :* les recommandations ont facilité la découverte de nouveaux lieux pertinents.
- *Qualité des suggestions :* les annotations proposées étaient satisfaisantes en termes de qualité.
- *Clarté :* les participants comprenaient les raisons pour lesquelles ces recommandations leur étaient faites.

– *Encouragement :* les recommandations ont incité à utiliser l'application de manière régulière.
– *Impact :* le module de recommandation a enrichi l'expérience utilisateur.

La seconde partie du questionnaire comportait des questions ouvertes visant à recueillir des informations qualitatives sur l'expérience des participants, abordant les aspects suivants :

– l'impact positif des recommandations, ainsi que les raisons expliquant cet impact ;
– les suggestions d'amélioration et d'adaptation des recommandations aux besoins individuels ;
– l'influence des recommandations sur l'utilisation de l'application et l'exploration urbaine.

4.3 Résultats

Les résultats sont présentés en trois parties, combinant données quantitatives et perceptions qualitatives : (1) l'impact global des notifications de recommandation sur l'activité des étudiants, (2) l'effet spécifique des incitations à l'annotation sur leur qualité (mesurée par les métriques dédiées), et (3) les retours directs des participants sur l'ensemble du système.

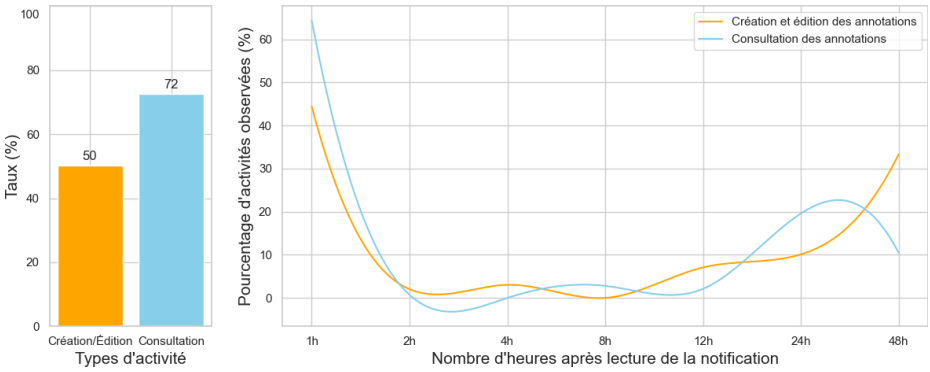


Fig. 3: Impact des recommandations sur l'activité des étudiants dans les 48h. À gauche : taux de notifications ayant suscité une activité. À droite : évolution des taux d'activité.

Impact sur les pratiques documentaires. L'analyse de l'impact des recommandations souligne leur rôle central dans l'engagement des étudiants, tant pour la création/modification d'annotations que pour la consultation des contenus partagés (Figure 3). Sur 197 notifications envoyées, 50,25% ont conduit à une

activité de création ou modification, et 72,59% à une interaction avec les annotations communautaires. Parmi les 23 étudiants ayant reçu des notifications, 19 (82,61%) ont créé ou modifié des annotations, et 20 (86,96%) ont consulté des contenus partagés. Ces niveaux de réceptivité indiquent que les recommandations sont perçues comme pertinentes et incitent à l'action.

L'analyse temporelle révèle une forte réactivité dans l'heure suivant la réception des notifications : 44,44% des créations/modifications et 64,34% des consultations s'y produisent. Cependant, l'engagement diminue ensuite rapidement (2,02% des créations/modifications et 0,70% des consultations entre 1 et 2 heures), devenant quasi inexistant au-delà de 2 heures. Malgré cette chute, un regain d'engagement pour les créations/modifications (33,33% des activités) est observé entre 24 et 48 heures. L'intégralité des consultations cumulées se produit dans les deux jours suivant les notifications, soulignant une dynamique d'interaction différée chez certains étudiants.

Ces résultats illustrent l'efficacité des recommandations pour générer un engagement rapide et significatif, tout en soulignant l'importance de stratégies adaptées. Une approche combinant incitations immédiates (pour capitaliser sur la réactivité initiale) et rappels différés (pour encourager des interactions prolongées) pourrait maximiser leur impact. Cela soutiendrait l'engagement dans la création/modification d'annotations et renforcerait l'intérêt pour la consultation des contributions communautaires.

Effets sur la qualité des annotations. Nous avons évalué l'impact des recommandations d'incitation à l'annotation sur leur qualité en analysant l'évolution hebdomadaire des métriques associées pour les utilisateurs ayant reçu des recommandations basées sur celles-ci. Les statistiques de ces évolutions sont présentées dans le Tableau 1 et la Figure 4.

Tableau 1: Nombre de recommandations envoyées, et évolution moyenne et médiane par stratégie.

Stratégie (métriques)	*Nbre*	*Taux d'évolution*	
		Médian	*Moyen*
Activité ($DocEng$)	24	7.5%	10.1%
Richesse Textuelle ($TextRich$)	20	5.9%	21%
Expression Émotionnelle ($EmoExp$)	21	5%	-5.9%
Richesse Graphique ($GraphRich$)	5	55%	55%
Icônes d'activité ($IconDiv_{\text{activity}}$)	5	8.9%	8.9%
Icônes environnementales ($IconDiv_{\text{env}}$)	7	5%	30%
Icônes sensorielles ($IconDiv_{\text{senses}}$)	9	16.5%	24.3%
Icônes sociales ($IconDiv_{\text{social}}$)	6	10.8%	15.3%
Icônes affectives ($IconDiv_{\text{affective}}$)	8	5%	19.6%

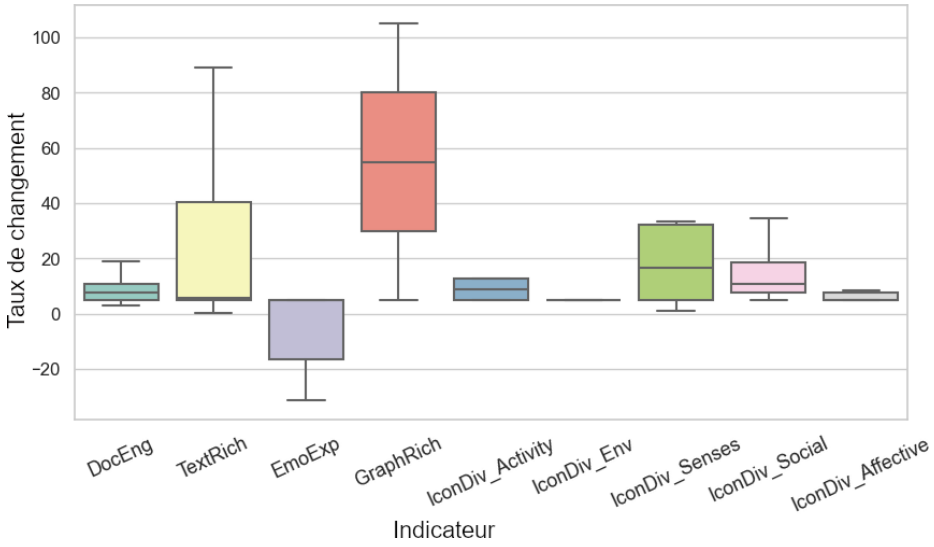


Fig. 4: Distribution de l'impact par type de recommandation.

Impact par stratégie d'incitation et perspectives des participants. Cette section examine l'impact des incitations à l'annotation *via* l'évolution hebdomadaire des métriques (Tableau 1, Figure 4). Les résultats sont complétés par les retours des étudiants lors des *focus groups* pour chaque aspect évalué.

Engagement dans l'activité de documentation. La métrique *DocEng*, qui mesure la création et l'édition d'annotations et de parcours, a augmenté en moyenne de 10,11% par semaine (médiane 7,50%) après 24 recommandations. Ces résultats indiquent une stimulation notable de l'engagement, surtout en début d'adoption. Les participants ont confirmé cette tendance, un d'entre eux déclarant : " *Les notifications n'étaient pas seulement des rappels, mais aussi des encouragements qui m'ont permis de rester engagé, surtout au début.* "

Richesse des récits textuels. La métrique *Richesse Textuelle (TextRich)*, évaluant la profondeur des annotations, a montré une hausse hebdomadaire médiane de 5,86% (moyenne 21%) après 20 recommandations. Celles-ci ont incité à des annotations plus détaillées. Un participant a souligné : " *Les suggestions d'annotation m'ont présenté des exemples bien rédigés qui m'ont inspiré, même si je ne me sens pas à la hauteur pour écrire à ce niveau.* " Ce commentaire illustre que, même pour ceux moins confiants, les recommandations ont favorisé une réflexion approfondie et une meilleure qualité de contenu.

Expression émotionnelle. L'évolution de la métrique *EmoExp*, qui suit l'utilisation d'icônes, de tags et d'émoticônes, a révélé une dynamique contrastée : une augmentation médiane hebdomadaire de 5% a été observée, tandis que la

moyenne a diminué de -5,96%. Certains participants ont trouvé ces incitations bénéfiques: "*Les icônes m'ont aidé à partager des émotions que j'ai souvent du mal à exprimer par mots.* " Cependant, d'autres ont jugé ces suggestions intrusives : "*Les suggestions m'ont poussé à m'exprimer d'une manière qui ne correspond pas à ma façon naturelle de décrire mes expériences.*" Ces retours mettent en lumière la nécessité d'une personnalisation accrue des recommandations.

Intégration d'éléments visuels. La *Richesse Graphique (GraphRich)* a montré une croissance hebdomadaire moyenne de 55% après seulement cinq recommandations, reflétant un usage déjà élevé d'éléments visuels. Un participant a commenté : " *Inclure des photos a donné vie à mes entrées, et les incitations m'ont rappelé de capturer des moments que j'aurais manqués.* " Les recommandations ont donc encouragé une documentation visuelle plus réfléchie et complète.

Diversification des éléments iconographiques. L'utilisation des icônes a significativement augmenté, avec des hausses moyennes de 8,85 % pour les icônes d'activité, 30 % pour les icônes de lieu, 24,27 % pour les icônes sensorielles, 15,27 % pour les icônes sociales, et 19,95 % pour les icônes affectives. Ces résultats sont corroborés par ce retour d'un participant : " *Les icônes m'ont aidé à exprimer mes expériences de manière plus nuancée, et je n'avais jamais réalisé à quel point elles pouvaient être puissantes.* " Ces tendances soulignent l'efficacité des recommandations pour enrichir et diversifier les annotations grâce à une utilisation créative des icônes.

Perception des participants sur le système de recommandations. Les perceptions des participants ont été recueillies *via* le questionnaire post-étude. Les avis étaient partagés sur la pertinence, la fréquence et l'impact des recommandations.

Tableau 2: Résumé des réponses aux questions fermées (25 participants).

Question	*Médiane*	*Moyenne (écart-type)*
Pertinence des recommandations	3.5	3.50 (1.21)
Variété des recommandations	4	3.66 (0.89)
Facilitation de la découverte	4	3.69 (0.92)
Qualité des suggestions	4	3.84 (1.04)
Clarté de la logique	3	3.38 (1.02)
Encouragement à l'engagement	4	3.61 (1.10)
Impact sur l'expérience	4	3.73 (1.96)

Le questionnaire à questions fermées (N=25) a permis de recueillir des informations sur la satisfaction concernant les aspects clés du système de recommandations personnalisées, avec une échelle de Likert de 1 (*tout à fait en désaccord*)

à 5 (*tout à fait d'accord*). Les résultats (Tableau 2) révèlent des perceptions partagées, soulignant points forts et axes d'amélioration.

La pertinence des recommandations a suscité une satisfaction modérée (médiane et moyenne de 3,5). Si certains les ont jugées utiles, d'autres les ont trouvées moins pertinentes, indiquant une réponse partielle aux attentes. La diversité a été bien accueillie (médiane 4, moyenne 3,66), bien que quelques participants aient noté une répétitivité à long terme, suggérant un besoin d'enrichissement pour maintenir l'engagement.

Concernant la découverte de lieux, les résultats sont positifs (médiane 4, moyenne 3,69). Les participants ont apprécié la capacité du système à leur faire découvrir des endroits, bien que certains, déjà familiers avec des zones, aient exprimé des frustrations, suggérant un besoin de recommandations plus avancées pour utilisateurs expérimentés.

Les suggestions personnalisées ont été bien perçues (médiane 4, moyenne 3,84), mais des améliorations sont nécessaires pour mieux répondre aux préférences spécifiques des étudiants. La compréhension de la logique de recommandation a été partagée (médiane 3, moyenne 3,38), et une meilleure explication de cette logique pourrait renforcer la confiance des étudiants et encourager leur engagement à long terme.

L'incitation à interagir avec l'application a montré une motivation modérée (médiane 4, moyenne 3,61). Si certains ont trouvé les incitations efficaces, d'autres les ont jugées moins pertinentes. Une approche plus personnalisée pourrait accroître l'engagement.

La qualité générale de l'expérience utilisateur a été perçue positivement (médiane 4, moyenne 3,73), mais certains ont exprimé une neutralité, suggérant des améliorations concernant par exemple l'interface et certaines fonctionnalités de l'application. Les résultats montrent une appréciation générale du système de recommandations, mais soulignent des axes d'amélioration clairs : personnalisation, transparence et diversité.

Les réponses aux questions ouvertes ont fourni des informations qualitatives supplémentaires. Les recommandations d'annotations ont facilité une documentation plus précise, mais plusieurs participants ont exprimé le besoin de comprendre les raisons des recommandations, ce qui pourrait renforcer confiance et engagement. La personnalisation selon les comportements passés et préférences individuelles est jugée essentielle pour améliorer la pertinence et fluidifier l'expérience.

5 Discussion et conclusion

Cette étude a examiné l'impact de l'intégration d'un système de recommandations personnalisées à une application mobile de documentation des expériences urbaines par les étudiants internationaux. Nos résultats indiquent globalement une influence positive du système sur l'engagement des étudiants et la richesse de leurs annotations. Toutefois, l'hétérogénéité observée dans l'impact de ces recommandations, variable en fonction des profils et des contextes individuels,

met en exergue la nécessité d'une approche plus nuancée en matière de personnalisation et d'adaptabilité.

L'analyse révèle que les recommandations ont favorisé un engagement accru des étudiants, se traduisant par une augmentation de la création et de la consultation de contenu, ainsi que par une diversification des formats utilisés (textes, images, vidéos). Cette multiplicité des supports a permis une expression plus riche des émotions et des réflexions, contribuant ainsi à l'enrichissement des annotations. Cependant, l'impact de ces recommandations s'est avéré inégal, certains étudiants exprimant des décalages entre les suggestions proposées et leurs attentes personnelles. Ce constat souligne l'importance d'une personnalisation plus fine, capable de prendre en compte la diversité des profils et des styles d'expression.

Un autre résultat notable est la diversification des pratiques documentaires. Le système a encouragé une exploration plus variée des formes d'expression, favorisant une expérience de documentation plus personnalisée. Néanmoins, cela a également révélé des défis d'accessibilité : si l'intégration des éléments visuels a été appréciée par certains, d'autres, moins familiarisés avec ces outils, ont rencontré des difficultés. Cette observation met en lumière l'importance de concilier créativité et accessibilité, en tenant compte de la diversité des compétences technologiques des étudiants.

L'incitation à documenter les expériences émotionnelles a permis à certains étudiants de dépasser la description factuelle pour des réflexions plus personnelles et critiques. Bien que cette profondeur de réflexion soit bénéfique, elle soulève également des interrogations quant au risque d'homogénéisation des réponses, qui pourrait nuire à la diversité des perspectives. L'enjeu est de soutenir une réflexivité authentique sans imposer des normes narratives trop rigides.

L'efficacité de la personnalisation des recommandations reste un défi majeur. Bien que le système ait globalement stimulé la réflexion, son impact a été limité par sa capacité d'adaptation aux besoins individuels et contextes variés. La diversité des pratiques documentaires et styles d'expression n'a pas été suffisamment prise en compte, révélant le besoin de solutions plus avancées (modèles adaptatifs, contextuels). Une plus grande transparence dans le processus de recommandation pourrait par ailleurs renforcer la confiance des étudiants et stimuler leur engagement à long terme.

Nos résultats suggèrent aussi que les recommandations ont facilité l'exploration de lieux inconnus, favorisant l'intégration dans un nouvel environnement culturel. Cependant, l'impact sur la socialisation et les interactions inter-étudiants, non pleinement exploré ici, constitue une piste pour des recherches futures.

Cette étude présente certaines limites qu'il convient de souligner. Premièrement, la taille modeste de l'échantillon (N=31) et l'absence d'une analyse longitudinale restreignent la généralisabilité des résultats et l'évaluation de l'impact à long terme. Deuxièmement, bien que l'origine géographique des participants ait été collectée, la taille et la répartition déséquilibrée de l'échantillon n'ont pas permis une analyse statistiquement fiable de l'influence potentielle des facteurs culturels, une dimension dont l'étude approfondie reste une perspective impor-

tante pour les travaux futurs. De même, la personnalisation, objectif central, n'a été que partiellement réalisée. Il est nécessaire de mieux équilibrer recommandations génériques et personnalisées, potentiellement *via* des algorithmes adaptatifs. De futures recherches pourraient étudier l'impact sur la socialisation, les pratiques linguistiques et la découverte urbaine, en veillant à disposer d'échantillons plus larges et diversifiés permettant d'intégrer l'analyse des facteurs culturels et contextuels, afin d'affiner le système et mieux comprendre son rôle dans l'intégration des étudiants internationaux.

En conclusion, cette étude confirme le potentiel des systèmes de recommandations personnalisées pour améliorer l'engagement étudiant, diversifier leurs pratiques documentaires et affiner la qualité de leurs annotations. Pour maximiser leur efficacité, il est crucial de renforcer la personnalisation, en tenant compte de la diversité des profils et contextes culturels. L'efficacité du système reposera sur sa capacité à évoluer, s'adapter continuellement aux besoins variés et intégrer les retours d'expérience. La prochaine étape sera d'étudier plus en profondeur l'impact à long terme de ces systèmes sur l'intégration sociale, le bien-être et la réussite académique des étudiants, tout en développant des solutions plus dynamiques, adaptatives et contextuelles.

Remerciements

Ce travail est soutenu par le projet MOBILES (ANR-20-CE38-0009) financé par l'Agence Nationale de la Recherche (ANR).

References

1. Chang, S., Gomes, C.: Digital journeys: A perspective on understanding the digital experiences of international students. Journal of International Students **7**(2), 347–466 (2017)
2. De Stefani, E., De Marco, D.: Language, gesture, and emotional communication: An embodied view of social interaction. Frontiers in Psychology **10**, 2063 (2019)
3. Drachsler, H., Verbert, K., Santos, O.C., Manouselis, N.: Panorama of recommender systems to support learning. Recommender systems handbook pp. 421–451 (2015)
4. Glass, C.R., Westmont, C.M.: Comparative effects of belongingness on the academic success and cross-cultural interactions of domestic and international students. International Journal of Intercultural Relations **38**, 106–119 (2014)
5. Griffin, M.: Visual communication. In: The Handbook of Communication History, pp. 133–152. Routledge (2013)
6. Guo, D., McTigue, E.M., Matthews, S.D., Zimmer, W.: The impact of visual displays on learning across the disciplines: A systematic review. Educational Psychology Review **32**(3), 627–656 (2020)
7. Kukulska-Hulme, A.: Mobile and personal learning for newcomers to a city. Electronic Journal of Foreign Language Teaching **17**(1), 93–103 (2020)
8. Lefevre, M., Benedetto, R., Sadallah, M., Chevalier, D., Cunty, C., Grassin, J.F., Joliveau, T., Mathian, H., Scheffler, C.: MOBILES: A Mobile Application for Documenting the Urban and Social Experiences of International Students. Software, Version 1.2.41 (2024), https://hal.science/hal-04771100

9. Loewen, S., Isbell, D.R., Sporn, Z.: The effectiveness of app-based language instruction for developing receptive linguistic knowledge and oral communicative ability. Foreign Language Annals **53**(2), 209–233 (2020)
10. Martirosyan, N.M., Bustamante, R., Saxon, D.P.: Academic and social support services for international students: Current practices. Journal of International Students **9**(1), 172–191 (2019)
11. Mueller, C.M., Richardson, P.: Talking about oneself: Effects of personalized prompts on l2 learners' speech. Journal of Second Language Studies **5**(2), 344–363 (2022)
12. Patel, N., Calhoun, D.W., Tolman, S.: Understanding the role of cultural competence in peer mentorship programs for international students: A student development theory perspective. Georgia Journal of College Student Affairs **40**(1), 62–80 (2024)
13. Poyrazli, S., Grahame, K.M.: Barriers to adjustment: Needs of international students within a semi-urban campus community. Journal of Instructional Psychology **34**(1), 28–45 (2007)
14. Sadallah, M., Lefevre, M.: MOBILES RecSys: A Personalized Recommendation System for Enhancing the Urban and Social Experiences of International Students. Software, Version 1.0 (2024), https://hal.science/hal-04771226
15. Sassi, I.B., Mellouli, S., Yahia, S.B.: Context-aware recommender systems in mobile environment: On the road of future research. Information Systems **72**, 27–61 (2017)
16. Settouti, L.S., Prié, Y., Cram, D., Champin, P.A., Mille, A.: A trace-based framework for supporting digital object memories. In: Workshops Proceedings of the 5th International Conference on Intelligent Environments. pp. 39–44. IOS Press (2009)
17. Silva, F.L., Slodkowski, B.K., Araújo da Silva, K.K., Cazella, S.C.: A systematic literature review on educational recommender systems for teaching and learning: research trends, limitations and opportunities. Education and Information Technologies **28**, 3289–3328 (2022)
18. Zhang, W., Chen, J.: Digital platforms for international student integration: A systematic literature review. Journal of International Students **12**(2), 405–425 (2022)
19. Zhou, Y., Jindal-Snape, D., Topping, K., Todman, J.: Theoretical models of culture shock and adaptation in international students in higher education. Studies in Higher Education **33**(1), 63–75 (2008)

Intégration responsable de l'IA Générative dans l'Éducation : proposition d'un plan d'actions stratégiques dirigé par les risques liés aux questions éthiques

Rémi Saurel[1,2][0009−0003−7969−1018], Franck Silvestre[1][0000−0002−1134−8200], Jean-Baptiste Raclet[1][0000−0001−7357−912X] et Emmanuel Lescure[2]

[1] IRIT, Université de Toulouse
[2] Kosmos Education

Résumé. Les résultats des premières études sur l'intégration de l'Intelligence Artificielle Générative (IAG) dans l'Éducation permettent de dégager 3 tendances pour les pays du *Global North* : les usages se développent très vite et de manière significative, mais avec un décalage croissant des usages en faveur des élèves, qui utilisent l'IAG dans un cadre non-contrôlé par les enseignants et pas toujours par les parents. Dans ce papier de positionnement, nous proposons un cadre d'actions stratégiques issu d'une analyse des risques liées aux questions éthiques identifiées par l'UNESCO. Notre approche vise à limiter les risques à criticité élevée afin d'augmenter les chances d'intégration responsable de l'IAG dans l'Éducation. Nous illustrons notre propos par la description du projet Konsolidation Dashboard (KDo), une initiative de recherche visant à informer et former sur les questions de l'intégration de l'IAG au sein de Tableaux de Bord d'Apprentissage dédiés aux enseignants.

Mots-clé: Intelligence artificielle générative · Éducation · Éthique · Analyse de risques · Littératie numérique

Abstract The results of the first studies on the integration of Generative Artificial Intelligence (GenAI) in education reveal 3 trends for the Global North countries: uses are developing very quickly and significantly, but with a growing shift in the use of GenAI in favour of learners who use GenAI in a context that is not controlled by teachers and not always by parents. In this position paper, we propose a framework for strategic actions based on an analysis of the risks associated with ethical issues identified by UNESCO. Our approach aims to limit high-criticality risks in order to increase the chances of responsibly integrating GenAI into education. We illustrate our position by describing the Konsolidation Dashboard (KDo) project, a research initiative aimed at informing and training on the issues involved in integrating GenAI into Learning Analytics Dashboards for teachers.

Keywords Generative Artificial Intelligence · Education · Ethic · Risk analysis · Digital literacy.

1 Introduction

L'intégration des technologies d'Intelligence Artificielle Générative (IAG) dans l'éducation suscite un intérêt croissant à l'échelle mondiale. Cette adoption rapide reflète les avancées technologiques majeures et leur potentiel à transformer les pratiques éducatives. Cependant, l'introduction de ces outils ne s'est pas faite de manière homogène, en particulier entre les principaux acteurs de la communauté éducative. Les études menées au Royaume-Uni [48] et en France [47] révèlent que les élèves mobilisent ces outils de manière plus importante que les enseignants, avec des disparités selon les niveaux d'éducation (secondaire vs. supérieur) et les contextes géographiques. Ces différences d ' adoption soulèvent des questions quant à la manière dont ces technologies s'intègrent dans des cadres éducatifs variés.

Cependant, les analyses actuelles sur le sujet présentent des limites importantes. Les études disponibles se concentrent principalement sur le supérieur et sur un outil particulier (ChatGPT) [54, 8]. Elles manquent également d'un cadre théorique pour analyser les politiques éducatives et se focalisent quasi exclusivement sur les pays du « Global North » — des régions développées comme l'Europe, l'Amérique du Nord et quelques pays d'Océanie [35, 21]. Cette concentration géographique limite la compréhension de l'impact global de l'IAG, tandis que l'évolution rapide des technologies et des usages invite à la prudence quant à l'interprétation des résultats, y compris ceux des études récentes.

Même avec ces limites, toutes les recherches que nous avons identifiées convergent sur certains aspects comme l'évolution rapide des technologies, des usages et du décalage d'utilisation entre élèves et enseignants. L'intégration de l'IAG, notamment du côté des élèves, se fait souvent de manière non pertinente pour l'amélioration des apprentissages. Du côté des enseignants, l'engagement plus modéré est attribué à un manque de compétences et de connaissances sur ces outils et leurs usages éducatifs.

Les résultats des premières études ne permettent pas de formuler des constats généralisables, mais sont suffisamment alarmants pour nous inviter à proposer des stratégies d'intégration de l'IAG dans l'éducation, guidées par une analyse de risques.

Nous nous intéressons dans ce papier de positionnement à la manière de limiter les risques à criticité élevée liés à l'intégration de l'IAG dans l'éducation.

La suite de cet article est composée de 4 sections. La section 2 présente les résultats révélés par les quelques études réalisées sur l'intégration de l'IAG dans l'éducation. La section 3 propose une analyse des risques s'appuyant sur les constats émergents des études à notre disposition et sur les dimensions éthiques documentées par l'UNESCO [50]. La section 4 définit un plan de gestion des risques incluant l'ensemble des parties prenantes du secteur éducatif. Enfin, la section 5 présente le projet *KDo* (Konsolidation Dashboard), un exemple concret d'initiative de recherche servant à informer et former sur les questions de l'intégration de l'IAG au sein de TBA (Tableaux de Bord d'Apprentissage) dédiés aux enseignants. La conclusion du papier insiste sur le caractère non définitif de

la proposition inhérent au contexte d'évolutions rapides et aux incertitudes induites.

2 Intégration de l'IAG dans l'éducation

L'intégration de l'IAG dans l'éducation est un phénomène en pleine expansion, avec des implications importantes pour les acteurs du secteur. Les études actuelles présentent un certain nombre de limites mentionnées dans l'introduction. Ces limites sont catalysées par les évolutions rapides de la technologie, des usages autour de l'IAG et de son intégration d'un point de vue pédagogique. C'est pour cela que nous identifions deux aspects majeurs qui guideront cette section : la rapidité d'évolution technologique et la rapidité d'évolution des usages.

2.1 Une rapidité d'évolution technologique

L'arrivée de ChatGPT en novembre 2022 a bouleversé un ensemble de secteurs, dont l'éducation. Tandis que les premiers grands modèles de langage (*Large Language Model, LLM*) étaient principalement connus par la communauté de l'IA, ChatGPT (avec son modèle GPT-3.5 lancé en 2022) a démocratisé l'accès et l'utilisation de ces technologies par le grand public. La puissance de calcul utilisée pour pré-entraîner ces modèles a grandement augmenté, permettant des avancées significatives en termes de qualité des réponses générées. Cette rapidité d'évolution technologique est présente sur deux niveaux : la puissance de calcul et la diversité de LLM.

La puissance de calcul. L'un des moteurs principaux de l'accélération technologique dans le domaine des modèles de langage réside dans l'augmentation de la puissance de calcul disponible. Cette évolution a permis de pré-entraîner des modèles de plus en plus complexes et performants.

Comme le souligne l'UNESCO dans son rapport [50], en seulement trois ans, les modèles sont passés de 117 millions de paramètres (pour GPT-1 en 2019) à 175 milliards de paramètres (pour GPT-3 en 2020). Ce bond massif dans le nombre de paramètres illustre l'augmentation des capacités des modèles à générer des réponses précises, à capter des nuances linguistiques et à étendre leur compréhension contextuelle.

La quantité de données utilisées pour l'entraînement a également connu une expansion sans précédent. Par exemple, GPT-1 a été formé sur un ensemble de 40 Go de données, tandis que GPT-3 utilisait 17 000 Go. En ce qui concerne GPT-4, bien que les chiffres exacts n'aient pas été publiés officiellement, l'UNESCO mentionne que ce modèle aurait été entraîné sur des données atteignant 1 000 000 Go, avec 170 000 milliards de paramètres.

La diversité de LLM. Tandis que les premiers modèles de langage, tels que GPT-3, étaient généralistes, de plus en plus de modèles sont conçus pour des cas d'utilisation spécifiques.

OpenAI, par exemple, a développé une gamme variée de modèles adaptés à différents usages, allant de la génération de texte (GPT) à la capacité de réflexion poussée (o1) [19]. META a introduit LLaMA (Large Language Model Meta AI) [13] avec différents modèles plus ou moins exigeants en ressources, tandis que des entreprises comme Mistral se spécialisent dans des solutions plus ciblées, telles que Pixtral (modèle multimodal) [1] et Ministral (un modèle compact pour des appareils à faible puissance de calcul) [32]. Midjourney, de son côté, est devenu une référence dans la création d'images, avec des versions comme Midjourney V6 et Niji Model 6 qui répondent à des besoins plus spécifiques, tels que la création artistique [31].

Par ailleurs, ces modèles diffèrent également sur leur nombre de paramètres (ou poids). Certains, comme GPT-4, comportent des centaines de milliards de paramètres et sont destinés à une utilisation intensive par des organisations dotées d'une infrastructure puissante. D'autres, plus petits, visent à rendre l'IA accessible sur des appareils individuels ou dans des environnements avec des ressources limitées.

2.2 Une rapidité d'évolution des usages

Tandis que les modèles de langage évoluent rapidement, les usages de ces technologies se multiplient, en particulier dans le domaine de l'éducation.

Malgré une adoption rapide de l'IAG par les apprenants mais aussi les enseignants, le décalage d'usage entre ces deux acteurs est significatif. Selon le *National Literacy Trust* [36], au Royaume-Uni, le pourcentage de 13-18 ans utilisant des outils d'IAG est passé de 37% en 2023 à 77% en 2024. Du côté des enseignants, toujours au Royaume-Uni, cette évolution a été beaucoup moins importante passant de 31% à (seulement) 48%. Cet écart grandissant entre les élèves et les enseignants est également observé en France d'après le rapport du Sénat de 2024-2025 sur le thème de l'IA et Éducation [47].

À partir de ces éléments, nous identifions deux tendances émergentes au sein des pays du *Global North* :

Tendance émergente 1

Les usages se développent très vite et de manière significative, en particulier chez les élèves, mais aussi chez les enseignants.

Tendance émergente 2

Il existe un décalage croissant de l'usage de l'IAG entre les élèves et les enseignants en faveur des élèves.

D'après l'étude du gouvernement du Royaume-Uni, parmi les élèves utilisant l'IAG dans un contexte éducatif, celle-ci est principalement réservée aux devoirs (61%) et à l'apprentissage à la maison (40%) [48]. Cette utilisation en dehors du cadre scolaire, non-contrôlée par les enseignants, soulève des inquiétudes quant à la supervision de ces usages. De plus, toujours d'après cette étude, 30% des

parents d'élèves (de 7 à 13 ans) au Royaume-Uni ne sont pas au courant de l'utilisation de l'IAG par leurs enfants [48].

Tendance émergente 3

Les élèves utilisent l'IAG dans un cadre non-contrôlé par les enseignants et pas toujours par les parents.

Dans la suite du document, nous considérons que l'intégration *responsable* d'une technologie dans l'Éducation est une intégration qui fait progresser de manière probante les apprentissages tout en prenant en compte les questions éthiques posées par cette intégration.

Les trois tendances identifiées nous suggèrent qu'une intégration responsable de l'IAG dans l'Éducation ne va pas de soi : ni les enseignants, ni les élèves ne paraissent armés des connaissances et compétences requises. Nous estimons qu'un plan stratégique d'urgence doit être mis en œuvre pour augmenter les chances d'une intégration responsable de l'IAG dans l'Éducation. Face au degré d'incertitude induit par la rapidité des évolutions tant sur le volet technique que sur le volet des usages, nous proposons de définir ce plan stratégique dans une démarche dirigée par les risques.

3 Des risques de l'intégration de l'IAG dans l'éducation

Dans cette section, en nous inscrivant dans les processus de management des risques décrits dans le PMBOK (Project Management Body of Knowledge) [18], nous identifions et structurons dans un premier temps un ensemble de risques liés à l'intégration de l'IAG dans l'éducation en nous appuyant sur le rapport de l'UNESCO [50]. En suivant, nous proposons une analyse qualitative de ces risques afin d'en estimer la criticité en combinant l'évaluation de leur impact et leur probabilité d'occurrence.

Notre analyse n'englobe pas tous les risques potentiels associés à l'IAG, car la rapidité d'évolution technologique et des usages fait apparaître en permanence de nouveaux défis. Slattery et al. [45] ont travaillé sur le MIT AI Risk Repository, un projet qui vise à identifier et à mettre à jour régulièrement les nouveaux risques afin d'aider chercheurs, développeurs et les différentes parties prenantes dans le domaine de l'IA. À partir de leur analyse, plus de 1000 risques ont été identifiés à l'aide de 50 cadres de risques liés à l'IA. Ils proposent 2 taxonomies : une taxonomie causale (comment, quand et pourquoi ces risques se produisent) et une taxonomie de domaine (7 domaines et 23 sous-domaines) pour classer ces risques.

Sur la base de cet exemple, nous sommes conscients que notre analyse n'est pas exhaustive et que d'autres risques sont déjà présents, tels que ceux concernant la dimension environnementale, ou pourraient émerger à l'avenir. Par conséquent, l'objectif de ce document n'est pas d'être exhaustif, mais plutôt d'initier une approche pour traiter ces risques en évaluant leur criticité, et d'illustrer cette

approche sur un sous-ensemble de risques balisé par le rapport de l'UNESCO [50].

3.1 Identification des risques

À partir du rapport de l'UNESCO [50, p.38] et de ses 5 dimensions éthiques liées exclusivement à l'utilisation des technologies d'IAG dans l'éducation et des descriptions de ces dimensions, nous avons extrait 8 risques, notés de (R1) à (R5.2) dans le Tableau 1.

Parmi ces risques, nous pouvons noter l'aggravation des inégalités éducatives (R1) liée à la dimension éthique *Accès et équité*, induite par le risque d'accroissement des disparités dans l'accès aux technologies et ressources éducatives.

La dimension *Connexion humaine* est quant à elle menacée par la réduction des interactions sociales et des aspects socio-émotionnels (R2) avec l'usage intensif d'agents artificiels interagissant en langage naturel.

Les risques liés au développement intellectuel humain sont multiples, allant de la limitation de l'autonomie et de la capacité d'agir (R3.1) à l'affaiblissement du développement intellectuel à long terme (R3.2) induit notamment par la fourniture de solutions préconçues par l'IAG.

La dimension *Impact psychologique* soulève des inquiétudes quant à la dégradation du bien-être émotionnel (R4.1) en lien direct avec la réduction des aspects socio-émotionnels (R2). Le risque de manipulation et d'influence malveillante (R4.2) a déjà fait l'objet d'une étude portant sur les interactions entre des enfants et des agents d'IA [12] comme Alexa d'Amazon ou Google Home.

Enfin, la dimension *Préjugé et discrimination cachée* soulève des inquiétudes quant à la génération de préjugés (R5.1) et de nouvelles formes de discriminations avec les résultats incertains des modèles d'IAG avec les phénomènes de biais et d'hallucinations [30, 20].

3.2 Estimation de l'impact

D'un point de vue analyse de risque, nous associons à chacun des risques du Tableau 1 un impact fort, car ils conduisent tous à une altérité des caractéristiques importantes de l'être humain (interactions sociales, capacités intellectuelles, bien-être, etc.).

Afin d'évaluer la criticité de ces risques, nous nous intéressons dans la section suivante à l'estimation de leur probabilité d'occurrence.

3.3 Estimation de la probabilité d'occurrence

Notre démarche pour fournir une estimation de la probabilité d'occurrence a consisté à relever, dans les études à notre disposition, les occurrences avérées de ces risques. Nous avons estimé que la présence dans les premières études de certains risques était suffisante pour affecter une probabilité d'occurrence élevée à ces derniers. Notre analyse a porté sur les modalités d'usages par les élèves et les enseignants.

Dimension éthique	**ID**	**Risque**
Accès et équité	R1	Aggravation des inégalités éducatives
Connexion humaine	R2	Réduction des interactions entre humains et des aspects socio-émotionnels
Développement intellectuel humain	R3.1	Limitation de l'autonomie et de la capacité d'agir
	R3.2	À long terme, affaiblissement du développement intellectuel
Impact psychologique	R4.1	Dégradation du bien-être émotionnel
	R4.2	Manipulation et influence malveillante
Préjugé et discrimination cachée	R5.1	Génération de préjugés
	R5.2	Génération de nouvelles formes de discriminations

Tableau 1. 5 Dimensions éthiques et risques associés

Élèves. L'utilisation des outils d'IAG par les élèves est en forte croissance. Selon une enquête réalisée au Royaume-Uni, le pourcentage d'élèves entre 13 et 18 ans utilisant ces outils est passé de 37 à 77% entre 2023 et 2024 [36]. Cette adoption rapide se concentre principalement sur des activités dans un contexte hors-classe, notamment les devoirs à la maison. Ce contexte soulève des inquiétudes quant au contrôle et à la supervision de ces usages [16], renforcées par le fait que les parents ne sont pas toujours conscients de l'utilisation de ces outils par leurs enfants [48]. En lien avec la dimension *Accès et équité* et le risque R1, cette utilisation à la maison peut accentuer les inégalités entre les élèves, en fonction de leur accès à ces technologies. L'utilisation massive peut être associée à la dimension *Connexion humaine* et le risque R2 en raison de la réduction des interactions sociales et de l'impact sur le développement cognitif des élèves.

Une étude menée par Kabir et al. [23] sur l'utilisation d'IAG pour résoudre des questions de programmation informatique a mis en évidence une tendance frappante : malgré la capacité de discerner les réponses de ChatGPT, 35% des élèves préfèrent les réponses générées par l'IA. Cependant, parmi ces réponses, trois sur quatre étaient incorrectes. Des résultats en lien avec cette étude ont été proposés par Darvishi et al. [9] lors d'un essai contrôlé randomisé. Celle-ci montre que les élèves tendent à s'appuyer sur l'assistance de l'IAG plutôt que d'apprendre à partir de celle-ci. Ces situations révèlent et renforcent le manque de compétences critiques telles que l'analyse, la synthèse et l'autonomie dans l'apprentissage, en lien direct avec R3.1 sur la dimension *Développement intellectuel humain*.

L'étude de Druga et al. [12] montre que les enfants personnifient et traitent les agents comme des humains, les perçoivent comme « amicaux » et « de confiance

». Cette imitation de l'interaction peut avoir des effets psychologiques négatifs comme la manipulation et une influence malveillante, identifée par le risque R4.2 de la dimension *Impact psychologique*.

Enfin, des *Préjugés et discriminations cachées* (dimension des risques R5.1 et R5.2) peuvent être exacerbés par les biais et les hallucinations générés par les modèles d'IAG [30, 20], et ces effets sont particulièrement marqués si les élèves ne sont pas formés à les détecter.

Ces risques deviennent encore plus préoccupants lorsque les élèves utilisent ces outils en dehors du cadre scolaire, sans supervision des enseignants et parfois sans que les parents en soient informés. Les enseignants ont donc un rôle crucial à jouer pour accompagner les élèves dans l'utilisation de ces technologies.

Enseignants. Selon le *National Literacy Trust*, le pourcentage d'enseignants du Royaume-Uni utilisant l'IAG est passé de 31% à 47% entre 2023 et 2024 [36]. Malgré cette augmentation, elle reste largement inférieure à celle des élèves avec plus de 77% d'utilisation pour les 13-18 ans, toujours d'après les résultats du *National Literacy Trust*.

Le DfE (Department for Education) du Royaume-Uni a réalisé un sondage en avril 2023 auprès de 2034 enseignants et 502 chefs d'établissement du pays sur l'utilisation de l'IAG. 83% des enseignants et chefs d'établissement ont indiqué ne pas avoir utilisé d'IAG, représentant environ 2100 personnes. Les résultats montrent que pour 58% de cet échantillon, la principale raison est leur manque de connaissance sur l'IAG et les cas d'usages pédagogiques possibles. Autrement dit, les enseignants et chefs d'établissement ne disposent pas des connaissances et compétences nécessaires pour intégrer l'IAG dans leur enseignement.

La littératie numérique regroupe un ensemble de compétences requises par les individus pour utiliser les technologies de manière efficace et critique. Elle inclut la capacité à rechercher, évaluer, utiliser et créer de l'information en utilisant les technologies numériques. Dans le contexte de l'éducation, la littératie numérique est essentielle pour permettre aux enseignants de guider les apprenants dans l'utilisation des outils numériques de façon responsable [43].

En 2023, seulement sept pays (Chine, Espagne, Finlande, Géorgie, Qatar, Thaïlande et Turquie) avaient développé des ressources spécifiques pour accompagner les enseignants dans l'intégration de l'IA dans leurs pratiques pédagogiques [50]. Cela souligne un retard global dans la formation des enseignants en littératie en IA, malgré une augmentation rapide de l'utilisation de l'IAG par les élèves.

Des initiatives émergent néanmoins. Par exemple, l'UNESCO [52] ou le gouvernement du Royaume-Uni [49] proposent des guides pratiques. En France, ces initiatives proviennent d'universités (Université de Bordeaux, [53]), de laboratoire de recherche (INRIA) avec des formations dédiées [17], et des communautés de réflexion [11] pilotées par la DNE (Direction du Numérique pour l'Éducation). Ces efforts visent à fournir aux enseignants les connaissances et compétences nécessaires pour intégrer l'IAG dans leurs méthodologies pédagogiques. Cependant, ces ressources semblent insuffisantes. En France, dans un

rapport du Sénat, une illustration de cette lacune est la prise de position des syndicats : *« L'ensemble des syndicats insistent sur les besoins de formation très importants des enseignants et la relative impréparation des établissements scolaires »* [47].

En synthèse, les études à notre disposition font état d'évènements qui sont des occurrences directes des risques R2, R3.1 et R4.2. Les enseignants manquant actuellement, pour beaucoup d'entre eux, de littératie en IAG, ne peuvent contribuer que faiblement à la prévention de ces risques.

3.4 Une criticité élevée pour certains risques

En analyse de risques, une échelle à 4 niveaux est souvent utilisée pour évaluer l'impact (de faible à très important) et la probabilité d'occurrence (de très improbable à très probable). La criticité, combinant l'impact et la probabilité d'occurrence peut varier de très basse à très élevée. Nous synthétisons les évaluations des risques liés à l'intégration de l'IAG dans l'Éducation dans le Tableau 2.

ID	Libellé	Impact	Probabilité	Criticité
R1	Aggravation des inégalités éducatives	Très important	?	?
R2	Réduction des interactions entre humains et des aspects socio-émotionnels		Très probable	Très élevée
R3.1	Limitation de l'autonomie et de la capacité d'agir		Très probable	Très élevée
R3.2	À long terme, affaiblissement du développement intellectuel		?	?
R4.1	Dégradation du bien-être émotionnel		?	?
R4.2	Manipulation et influence malveillante		Très probable	Très élevée
R5.1	Génération de préjugés		?	?
R5.2	Génération de nouvelles formes de discriminations		?	?

Tableau 2. Impact, probabilité et criticité des risques

Dans les études identifiées, un manque de données probantes ne nous permet pas de déterminer la criticité pour 5 des risques identifiés. Cependant, les risques pour lesquels nous avons des éléments nous incitent à les associer à une criticité élevée. Nous proposons un plan de gestion se focalisant sur des actions stratégiques visant à limiter les risques à criticité élevée.

4 Proposition d'un plan de gestion de risques

4.1 Implication de l'ensemble des parties prenantes

La littérature sur les Environnements Informatiques pour l'Apprentissage Humain (EIAH) souligne l'importance d'une approche centrée sur l'humain [37, 14, 3, 56, 29]. Cette approche met en avant la nécessité de considérer les besoins et les attentes des utilisateurs finaux, en l'occurence les enseignants et les élèves, mais aussi des autres parties prenantes telles que les parents, les organisations en charge de la définition et de la mise en œuvre des politiques éducatives, les chercheurs, les développeurs et les éditeurs de logiciels dans le processus de conception et de déploiement des systèmes.

À travers une revue systématique, Alfredo et al. [3] ont proposé trois recommandations en liant les approches centrées humains et l'IA dans l'éducation:

- l'équilibre sur l'implication des parties prenantes dans la conception et le déploiement des systèmes ;
- l'implication des utilisateurs finaux pour définir l'équilibre entre le contrôle humain et l'automatisation ;
- la conception guidée par la sécurité et la fiabilité.

L'écosystème de l'éducation est complexe et nécessite une collaboration active entre les différents acteurs pour garantir une intégration pertinente de l'IAG dans les pratiques pédagogiques. Défini autour de 4 mots clés - soutenir, fournir, encadrer et informer - nous proposons un schéma de l'écosystème de l'éducation (voir Figure 1) qui met en lumière les interactions entre les différentes parties prenantes.

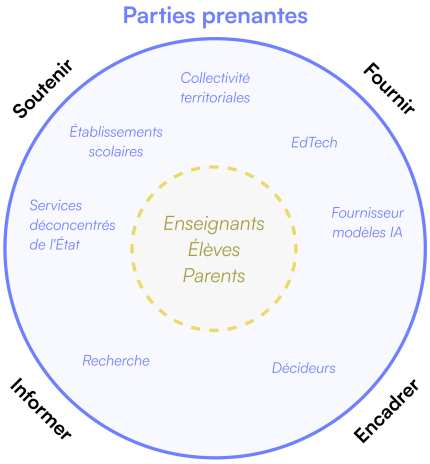


Fig. 1. Parties prenantes et leurs rôles dans l'écosystème de l'éducation

Le soutien, par l'intermédiaire des établissements scolaires et des services déconcentrés de l'État, comprend le suivi, l'accompagnement et la formation des enseignants, des élèves et des parents. Des EdTech (ici, les entreprises des technologies éducatives) et fournisseurs de modèles d'IA doivent fournir services et outils en cohérence avec le contexte (niveau et discipline). L'encadrement consiste à définir les objectifs pédagogiques et le cadre réglementaire, par exemple en produisant des législations et des guides de mise en œuvre comme cela a pu être le cas dans d'autres pays [50]. Enfin, informer l'ensemble des acteurs en mobilisant la recherche et la communication est essentiel pour une prise de décision éclairée.

Wang et al. [54] proposent l'idée que l'effort sur le développement des politiques autour de l'IA doit être partagé entre différentes parties prenantes comme les enseignants, mais aussi les organisations en charge de la définition et de la mise en œuvre des politiques éducatives ou les services déconcentrés de l'État.

La section suivante présente les actions stratégiques que nous proposons pour les parties prenantes jugées prioritaires.

4.2 Actions stratégiques pour cibles prioritaires

Nous identifions 3 cibles prioritaires (Figure 2), enseignants, élèves et parents, pour mener des actions stratégiques en vue d'augmenter les chances d'une intégration responsable de l'IAG dans l'éducation (Figure 1). Les actions au sein de chacun des axes visent la limitation des risques de criticité élevée.

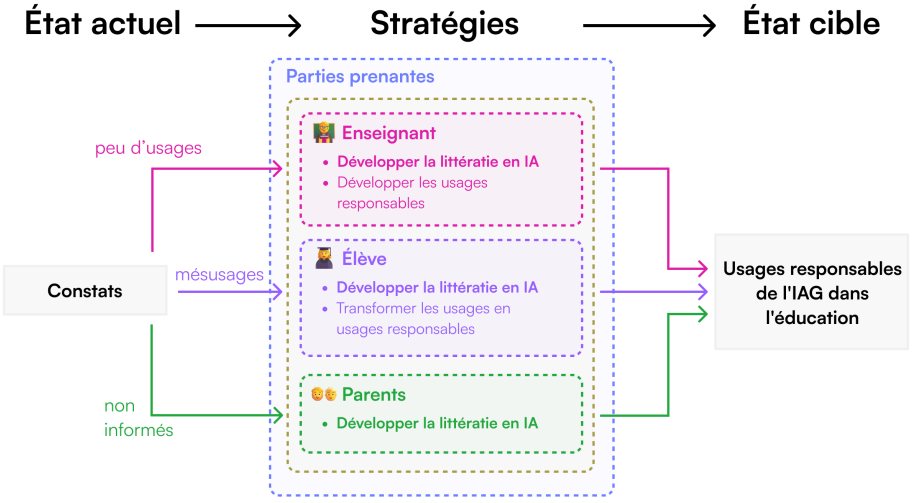


Fig. 2. Axes stratégiques

Le développement de la littératie en IA est une action stratégique commune aux 3 acteurs explicités dans la Figure 2. Long et Magerko [27] ont travaillé sur la définition de la littératie en IA en proposant un ensemble de compétences, tout comme l'UNESCO pour les enseignants [52] et les étudiants [51]. Ng et al. [37] ont conceptualisé cette littératie en utilisant des cadres théoriques issues de la psychologie avec la Taxonomie de Bloom [25] ou encore TPACK (*Technological Pedagogical Content Knowledge*) introduit par Graham [15]). Les résultats de l'étude qualitative de Ding et al. [10] vont dans ce sens : il est important de promouvoir la littératie en IA chez les enseignants pour leur mise en œuvre efficace et responsable des technologies d'IA dans les contextes éducatifs. Cette action de formation sur le développement de la littératie en IA doit permettre aux enseignants de soutenir les élèves dans l'utilisation de l'IAG de manière responsable.

Ainsi, la formation des enseignants à l'IAG n'est pas seulement une question de compétence technique : elle est également un levier stratégique pour préparer les jeunes générations à un monde où l'IA occupe une place centrale. C'est d'ailleurs ce que pense le Sénat français [47] ou les universités internationales [21, 46] : l'IA sera une compétence clé pour les métiers de demain. Un investissement massif dans la formation initiale et continue des enseignants apparaît donc comme une priorité pour tous les systèmes éducatifs.

Van Niekerk et al. [38] ont montré qu'une intervention éducative ciblée modifie significativement les comportements des élèves dans leur utilisation de l'IAG. Après cette formation, les élèves utilisent moins l'IAG pour des éléments cruciaux tels que la structuration, les preuves et les références, et adoptent une approche plus critique vis-à-vis des résultats générés. De plus, ils exploitent mieux les outils pour des activités comme le résumé ou l'adoption d'un ton formel, montrant une meilleure compréhension des forces et des limites de l'IAG. Ces résultats montrent que la formation des élèves à l'IAG est également un levier important pour réduire les risques liés à l'utilisation de ces technologies comme R3.1 (Limitation autonomie et de la capacité à agir) mais aussi R4.2 (Manipulation et influence malveillante) du Tableau 3.1.

Enfin, les parents jouent un rôle crucial dans l'accompagnement des élèves dans l'utilisation de l'IAG en limitant les risques R2 (Réduction des interactions entre humains et des aspects socio-émotionnels) mais aussi R4.2 (Manipulation et influence malveillante). L'utilisation de ces technologies majoritairement dans un contexte hors-classe souligne l'importance de sensibiliser les parents aux enjeux criants en développant une littératie en IA adaptée à leurs besoins.

Ces axes stratégiques ont besoin d'être développés et de s'appuyer sur des données probantes qui manquent cruellement en raison des différences de rythmes entre l'évolution des technologies et celle des pratiques pédagogiques.

Le besoin de données probantes amenées par des recherches est essentiel pour garantir des actions stratégiques efficaces.

4.3 Informer et former

L'effort sur la dimension *Informer* de la Figure 1 est urgent car on manque de données probantes sur ce qui fonctionne ou non en termes d'apprentissage quand on intègre l'IAG. Il est nécessaire de mener des recherches pour informer et former les acteurs du secteur éducatif sur l'utilisation responsable de l'IAG dans l'éducation. Cependant, cette formation doit se baser sur des travaux relativement récents et avec une rigueur méthodologique élevée pour être pertinente et efficace.

Les travaux de Yan et al. [55] insistent sur trois besoins essentiels pour une intégration réussie de l'IAG dans l'apprentissage humain :

- la littératie en IA;
- la prise de décision basée sur des preuves (ou Evidence-Based decision making);
- la rigueur méthodologique dans la recherche lors de l'utilisation de l'IAG.

Les deux derniers points concernent directement la recherche. Ils sont liés à la nécessité d'établir un cadre théorique robuste pour l'évaluation de la pertinence pédagogique des outils d'IAG. Malgré l'évolution rapide des technologies, la mise en place de cadres et de méthodes d'évaluation rigoureuses des processus, objectifs et expériences pédagogiques, incluant éventuellement des essais contrôlés randomisés à grande échelle, constitue une condition essentielle pour garantir des recherches de qualité permettant de lever une partie des incertitudes entourant l'intégration responsable de l'IAG dans le domaine éducatif.

La section suivante illustre par un exemple de projet de recherche comment celle-ci peut s'insérer dans notre plan de gestion des risques. Les enseignants représentent le premier facteur pouvant limiter les risques liés à l'intégration de l'IAG dans l'éducation. En effet, ils sont les premiers à être en contact avec les élèves et à les accompagner dans l'utilisation de ces technologies dans un cadre éducatif. Il est donc essentiel de démarrer les actions stratégiques envers eux. Le projet KDo s'inscrit dans cette démarche.

5 Le projet KDo

Dans le projet de recherche KDo nous souhaitons étudier comment le numérique et l'IA peuvent soutenir l'implication des enseignants dans la régulation du travail hors classe et pour quels bénéfices.

À travers la recherche et basée sur l'application de travail scolaire hors-classe Konsolidation [39, 40], KDo est un projet qui s'inscrit dans la continuité de la thèse d'Ollivier. En adoptant une approche *Design-Based Research* (DBR) [4, 28], nous plaçons les parties prenantes — enseignants, établissements scolaires et services déconcentrés de l'État (comme les académies) — au cœur du processus de recherche dès le début du projet. Cette implication des parties prenantes est renforcée par le dispositif CIFRE (Convention Industrielle de Formation par

la Recherche), qui permet d'associer étroitement l'EdTech Kosmos[3] dès les premières phases de la recherche.

L'étude qualitative d'Ollivier et al. [39, 40] a permis d'expliciter le besoin de certains enseignants de réguler le travail hors-classe à l'aide d'indicateurs actionnables. Ces indicateurs sont issus de l'analyse de l'apprentissage, également connu sous le nom de *Learning Analytics* (LA), qui consiste à collecter et à exploiter les données des apprenants afin de mieux comprendre et optimiser leur processus d'apprentissage [26]. Dans le champ des LA, les tableaux de bord d'apprentissage (TBA) [44] ont déjà montré des résultats intéressants grâce à des indicateurs actionnables [34] (qui permettent d'améliorer l'enseignement à l'aide d'actions clairement identifiables). Cela a été mis en évidence lors d'études sur les types de feedback proposés aux élèves en difficulté [6] ou l'aide à l'identification de ces élèves [42].

Cependant, une conception ou utilisation inadéquate des TBA peut rendre difficile l'analyse, conduire à de fausses conclusions ou être incompatible avec les exigences pédagogiques des enseignants [34, 7, 24]. Pour y remédier, un changement d'orientation est amorcé : des TBA exploratoires vers des TBA explicatifs [14, 41, 22].

Molenaar [33] décrit les interactions enseignant-IA via le concept de *Hybrid Human-AI Regulation* (HHAIR). L'IA, en déchargeant certaines tâches [5], permet aux enseignants de se focaliser sur des activités complexes et à forte valeur ajoutée. Toutefois, l'enseignant doit rester au centre du processus éducatif, en conservant sa « capacité à agir » (*agency*) pour maîtriser les décisions et actions [2].

Dans cette perspective, KDo s'inscrit dans le courant HHAIR en étudiant la mobilisation de l'IAG dans les processus éducatifs. Concrètement, nous souhaitons étudier comment l'IAG pourra être intégrée à l'analyse des productions des élèves afin de fournir des informations actionnables à l'enseignant, tout en permettant une meilleure compréhension des visualisations et des indicateurs pour soutenir une prise de décision éclairée. En offrant une assistance par l'IAG d'abord aux enseignants, le projet KDo s'inscrit dans le premier ensemble d'actions stratégiques : les actions ciblant les enseignants. Le TBA proposé n'a donc pas comme objectif de détecter des usages des IAG fait par les élèves mais vise, in fine, à améliorer les pratiques de révision des élèves notamment au regard des effets tests et espacement.

6 Conclusion

Les résultats des premières études sur l'intégration de l'Intelligence Artificielle Générative (IAG) dans l'Éducation permettent de dégager 3 tendances pour les pays du *Global North* : les usages se développent très vite et de manière significative, mais avec un décalage croissant des usages en faveur des élèves, qui utilisent l'IAG dans un cadre non-contrôlé par les enseignants et pas toujours par les parents.

[3] Kosmos Education

À travers une démarche dirigée par les risques et en nous appuyant sur le rapport de l'UNESCO et ses 5 dimensions éthiques, nous avons identifié 8 risques liés à l'intégration de l'IAG dans l'éducation. Après une estimation de l'impact et de la probabilité d'occurrences à partir des quelques études disponibles, nous avons identifié une criticité élevée pour certains de ces risques.

Afin de réduire cette criticité, nous avons proposé un plan de gestion des risques orienté vers 3 cibles prioritaires (enseignants, élèves et parents) et impliquant l'ensemble des parties prenantes de l'écosystème de l'éducation.

Le projet KDo illustre comment la recherche peut s'insérer dans ce plan de gestion des risques en s'appuyant sur une approche DBR pour étudier comment le numérique et l'IA peuvent soutenir l'implication des enseignants dans la régulation du travail hors classe et pour quels bénéfices.

References

1. Agrawal, P., Antoniak, S., Hanna, E.B., Bout, B., Chaplot, D., Chudnovsky, J., Costa, D., De Monicault, B., Garg, S., Gervet, T., et al.: Pixtral 12b. arXiv preprint arXiv:2410.07073 (2024)
2. Albert Bandura: Human agency in social cognitive theory (1989), https://pubmed.ncbi.nlm.nih.gov/2782727/
3. Alfredo, R., Echeverria, V., Yueqiao, J., Lixiang, Y., Swiecki, Z., Gašević, D., Martinez-Maldonado, R.: Human-centred learning analytics and AI in education: A systematic literature review. Computers and Education: Artificial Intelligence **6**, 100215 (Jun 2024). https://doi.org/10.1016/j.caeai.2024.100215, https://www.sciencedirect.com/science/article/pii/S2666920X2400016X
4. Anderson, T., Shattuck, J.: Design-based research: A decade of progress in education research? Educational researcher **41**(1), 16–25 (2012)
5. Buxbaum-Conradi, S., Redlich, T., Branding, J.H.: Conceptualizing Hybrid Human-Machine Systems and Interaction. In: 2016 49th Hawaii International Conference on System Sciences (HICSS). pp. 551–559 (Jan 2016). https://doi.org/10.1109/HICSS.2016.75, https://ieeexplore.ieee.org/document/7427251, iSSN: 1530-1605
6. Knoop-van Campen, C.A.N., Wise, A., Molenaar, I.: The equalizing effect of teacher dashboards on feedback in K-12 classrooms. Interactive Learning Environments **31**(6), 3447–3463 (Aug 2023). https://doi.org/10.1080/10494820.2021.1931346, https://www.tandfonline.com/doi/full/10.1080/10494820.2021.1931346
7. Campos, F.C., Ahn, J., DiGiacomo, D.K., Nguyen, H., Hays, M.: Making Sense of Sensemaking: Understanding How K–12 Teachers and Coaches React to Visual Analytics. Journal of Learning Analytics **8**(3), 60–80 (Jul 2021). https://doi.org/10.18608/jla.2021.7113, https://www.learning-analytics.info/index.php/JLA/article/view/7113, number: 3
8. Cheng, M.W.T., YIM, I.H.Y.: Examining the use of ChatGPT in public universities in Hong Kong: a case study of restricted access areas. Discover Education **3**(1), 1 (Jan 2024). https://doi.org/10.1007/s44217-023-00081-8, https://doi.org/10.1007/s44217-023-00081-8
9. Darvishi, A., Khosravi, H., Sadiq, S., Gašević, D., Siemens, G.: Impact of AI assistance on student agency. Computers & Education

210, 104967 (Mar 2024). https://doi.org/10.1016/j.compedu.2023.104967, https://www.sciencedirect.com/science/article/pii/S0360131523002440
10. Ding, A.C.E., Shi, L., Yang, H., Choi, I.: Enhancing teacher AI literacy and integration through different types of cases in teacher professional development. Computers and Education Open **6**, 100178 (Jun 2024). https://doi.org/10.1016/j.caeo.2024.100178, https://www.sciencedirect.com/science/article/pii/S2666557324000193
11. DNE-TN2: CREIA – Communauté de Réflexion en Éducation sur l'Intelligence Artificielle (Sep 2024). https://doi.org/10.58079/12cac, https://edunumrech.hypotheses.org/12979, iSSN: 2725-125X
12. Druga, S., Williams, R., Breazeal, C., Resnick, M.: "Hey Google is it OK if I eat you?": Initial Explorations in Child-Agent Interaction. In: Proceedings of the 2017 Conference on Interaction Design and Children. pp. 595–600. IDC '17, Association for Computing Machinery, New York, NY, USA (Jun 2017). https://doi.org/10.1145/3078072.3084330, https://dl.acm.org/doi/10.1145/3078072.3084330
13. Dubey, A., Jauhri, A., Pandey, A., Kadian, A., Al-Dahle, A., Letman, A., Mathur, A., Schelten, A., Yang, A., Fan, A., et al.: The llama 3 herd of models. arXiv preprint arXiv:2407.21783 (2024)
14. Echeverria, V., Martinez-Maldonado, R., Granda, R., Chiluiza, K., Conati, C., Buckingham Shum, S.: Driving data storytelling from learning design. In: Proceedings of the 8th International Conference on Learning Analytics and Knowledge. pp. 131–140. LAK '18, Association for Computing Machinery, New York, NY, USA (Mar 2018). https://doi.org/10.1145/3170358.3170380, https://doi.org/10.1145/3170358.3170380
15. Graham, C.R.: Theoretical considerations for understanding technological pedagogical content knowledge (TPACK). Computers & Education **57**(3), 1953–1960 (Nov 2011). https://doi.org/10.1016/j.compedu.2011.04.010, https://www.sciencedirect.com/science/article/pii/S0360131511000911
16. Hingley, S., Gaskell, E., Morris, J., Man, C., Thornton, A.: Parent, pupil and learner panel 22/23 April/May wave
17. INRIA: Intelligence artificielle pour et par les enseignants, https://www.fun-mooc.fr/fr/cours/intelligence-artificielle-pour-et-par-les-enseignants-ai4t/
18. Institute, P.M.: A guide to the Project Management Body of Knowledge (PMBOK guide). Project Management Institute (2021)
19. Jaech, A., Kalai, A., Lerer, A., Richardson, A., El-Kishky, A., Low, A., Helyar, A., Madry, A., Beutel, A., Carney, A., et al.: Openai o1 system card. arXiv preprint arXiv:2412.16720 (2024)
20. Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y., Chen, D., Dai, W., Chan, H.S., Madotto, A., Fung, P.: Survey of Hallucination in Natural Language Generation (Jul 2024). https://doi.org/10.48550/arXiv.2202.03629, http://arxiv.org/abs/2202.03629, arXiv:2202.03629
21. Jin, Y., Yan, L., Echeverria, V., Gašević, D., Martinez-Maldonado, R.: Generative AI in higher education: A global perspective of institutional adoption policies and guidelines. Computers and Education: Artificial Intelligence **8**, 100348 (Jun 2025). https://doi.org/10.1016/j.caeai.2024.100348, https://www.sciencedirect.com/science/article/pii/S2666920X24001516
22. Jin, Y., Yang, K., Yan, L., Echeverria, V., Zhao, L., Alfredo, R., Milesi, M., Fan, J., Li, X., Gašević, D., Martinez-Maldonado, R.: Chatting with a Learning Analytics Dashboard: The Role of Generative AI Literacy on

Learner Interaction with Conventional and Scaffolding Chatbots (Nov 2024). https://doi.org/10.48550/arXiv.2411.15597, http://arxiv.org/abs/2411.15597, arXiv:2411.15597 [cs]
23. Kabir, S., Udo-Imeh, D.N., Kou, B., Zhang, T.: Is Stack Overflow Obsolete? An Empirical Study of the Characteristics of ChatGPT Answers to Stack Overflow Questions (Feb 2024). https://doi.org/10.48550/arXiv.2308.02312, http://arxiv.org/abs/2308.02312, arXiv:2308.02312
24. Kaliisa, R., Mørch, A.I., Kluge, A.: My Point of Departure for Analytics is Extreme Skepticism': Implications Derived from An Investigation of University Teachers' Learning Analytics Perspectives and Design Practices. Technology, Knowledge and Learning **27**(2), 505–527 (Jun 2022). https://doi.org/10.1007/s10758-020-09488-w, https://doi.org/10.1007/s10758-020-09488-w
25. Krathwohl, D.R.: A Revision of Bloom's Taxonomy: An Overview. Theory Into Practice **41**(4), 212–218 (2002), https://www.jstor.org/stable/1477405, publisher: Taylor & Francis, Ltd.
26. Lang, C., Wise, A., Merceron, A., Gašević, D., Siemens, G.: What is learning analytics. In: Lang, C., Siemens, G., Wise, A.F. (eds.) The Handbook of Learning Analytics, pp. 8–18. SOLAR, 2 edn. (2022). https://doi.org/10.18608/hla22.001, https://www.solaresearch.org/publications/hla-22/hla22-chapter1/
27. Long, D., Magerko, B.: What is AI Literacy? Competencies and Design Considerations. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. pp. 1–16. CHI '20, Association for Computing Machinery, New York, NY, USA (Apr 2020). https://doi.org/10.1145/3313831.3376727, https://dl.acm.org/doi/10.1145/3313831.3376727
28. Mandran, N., Vermeulen, M., Prior, E.: Comment guider les doctorants dans l'utilisation du design-based research? In: 10e Conférence sur les Environnements Informatiques pour l'Apprentissage Humain. pp. 130–141 (2021)
29. Matcha, W., Uzir, N.A., Gašević, D., Pardo, A.: A Systematic Review of Empirical Studies on Learning Analytics Dashboards: A Self-Regulated Learning Perspective. IEEE Transactions on Learning Technologies **13**(2), 226–245 (Apr 2020). https://doi.org/10.1109/TLT.2019.2916802, https://ieeexplore.ieee.org/document/8713912, conference Name: IEEE Transactions on Learning Technologies
30. Maynez, J., Narayan, S., Bohnet, B., McDonald, R.: On Faithfulness and Factuality in Abstractive Summarization (May 2020). https://doi.org/10.48550/arXiv.2005.00661, http://arxiv.org/abs/2005.00661, arXiv:2005.00661
31. Midjourney: Midjourney Models, https://docs.midjourney.com/docs/model-versions
32. Mistral: Mistral Models, https://docs.mistral.ai/getting-started/models/models_overview/
33. Molenaar, I.: The concept of hybrid human-AI regulation: Exemplifying how to support young learners' self-regulated learning. Computers and Education: Artificial Intelligence **3**, 100070 (Jan 2022). https://doi.org/10.1016/j.caeai.2022.100070, https://www.sciencedirect.com/science/article/pii/S2666920X2200025X
34. Molenaar, I., Knoop-van Campen, C.A.N.: How Teachers Make Dashboard Information Actionable. IEEE Transactions on Learning Technologies **12**(3), 347–355 (Jul 2019). https://doi.org/10.1109/TLT.2018.2851585, https://ieeexplore.ieee.org/document/8400475, conference Name: IEEE Transactions on Learning Technologies

35. Moorhouse, B.L., Yeo, M.A., Wan, Y.: Generative AI tools and assessment: Guidelines of the world's top-ranking universities. Computers and Education Open **5**, 100151 (Dec 2023). https://doi.org/10.1016/j.caeo.2023.100151, https://www.sciencedirect.com/science/article/pii/S2666557323000290
36. National Literacy Trust: Children, young people and teachers' use of generative AI to support literacy in 2024 (Jun 2024), https://literacytrust.org.uk/research-services/research-reports/children-young-people-and-teachers-use-of-generative-ai-to-support-literacy-in-2024/
37. Ng, D.T.K., Leung, J.K.L., Chu, S.K.W., Qiao, M.S.: Conceptualizing AI literacy: An exploratory review. Computers and Education: Artificial Intelligence **2**, 100041 (Jan 2021). https://doi.org/10.1016/j.caeai.2021.100041, https://www.sciencedirect.com/science/article/pii/S2666920X21000357
38. van Niekerk, J., Delport, P.M.J., Sutherland, I.: Addressing the use of generative AI in academic writing. Computers and Education: Artificial Intelligence **8**, 100342 (Jun 2025). https://doi.org/10.1016/j.caeai.2024.100342, https://www.sciencedirect.com/science/article/pii/S2666920X24001450
39. Ollivier, D., Silvestre, F., Raclet, J.B., Lescure, E., Broisin, J.: Conception d'un système de révisions: une étude qualitative exploratoire pour identifier les besoins des enseignants et des élèves. In: 11ème Conférence sur les Environnements Informatiques pour l'Apprentissage Humain (EIAH 2023). pp. 36–41 (2023)
40. Ollivier, D., Silvestre, F., Raclet, J.B., Lescure, E., Broisin, J.: Designing a revision system: An exploratory qualitative study to identify the needs of french teachers and students. In: European Conference on Technology Enhanced Learning. pp. 294–307. Springer (2023)
41. Pozdniakov, S., Martinez-Maldonado, R., Tsai, Y.S., Echeverria, V., Srivastava, N., Gasevic, D.: How Do Teachers Use Dashboards Enhanced with Data Storytelling Elements According to their Data Visualisation Literacy Skills? In: LAK23: 13th International Learning Analytics and Knowledge Conference. pp. 89–99. LAK2023, Association for Computing Machinery, New York, NY, USA (Mar 2023). https://doi.org/10.1145/3576050.3576063, https://doi.org/10.1145/3576050.3576063
42. Raclet, J.B., Silvestre, F.: Git4school: A dashboard for supporting teacher interventions in software engineering courses. In: European Conference on Technology Enhanced Learning. pp. 392–397. Springer (2020)
43. Reddy, P., Sharma, B., Chaudhary, K.: Digital Literacy: A Review of Literature. International Journal of Technoethics (IJT) **11**(2), 65–94 (Jul 2020). https://doi.org/10.4018/IJT.20200701.oa1, https://www.igi-global.com/article/digital-literacy/www.igi-global.com/article/digital-literacy/258971, publisher: IGI Global Scientific Publishing
44. Schwendimann, B.A., Rodríguez-Triana, M.J., Vozniuk, A., Prieto, L.P., Boroujeni, M.S., Holzer, A., Gillet, D., Dillenbourg, P.: Perceiving Learning at a Glance: A Systematic Literature Review of Learning Dashboard Research. IEEE Transactions on Learning Technologies **10**(1), 30–41 (Jan 2017). https://doi.org/10.1109/TLT.2016.2599522, https://ieeexplore.ieee.org/document/7542151, conference Name: IEEE Transactions on Learning Technologies
45. Slattery, P., Saeri, A.K., Grundy, E.A., Graham, J., Noetel, M., Uuk, R., Dao, J., Pour, S., Casper, S., Thompson, N.: The ai risk repository: A comprehensive meta-review, database, and taxonomy of risks from artificial intelligence. arXiv preprint arXiv:2408.12622 (2024)

46. Sullivan, M., Kelly, A., McLaughlan, P.: ChatGPT in higher education: Considerations for academic integrity and student learning. Journal of Applied Learning & Teaching **6** (Mar 2023). https://doi.org/10.37074/jalt.2023.6.1.17
47. Sénat, Gouvernement Français: IA et éducation, https://www.senat.fr/rap/r24-101/r24-101.html
48. UK Government [Social Research]: Generative AI in education: educator and expert views (2024), https://www.gov.uk/government/publications/generative-ai-in-education-educator-and-expert-views
49. UK Government [Social Research]: Use Cases for Generative AI in Education: User Research Report (Aug 2024), https://shorturl.at/OFcoU
50. UNESCO: Guidance for generative AI in education and research. UNESCO (2023), https://www.unesco.org/en/articles/guidance-generative-ai-education-and-research
51. UNESCO: AI competency framework for students (2024), https://www.unesco.org/en/articles/ai-competency-framework-students
52. UNESCO: AI competency framework for teachers (2024), https://www.unesco.org/en/articles/ai-competency-framework-teachers
53. Université de Bordeaux: Activités de pédagogie active pour faire utiliser l'IA (2024), https://enseigner.u-bordeaux.fr/application/files/5617/0773/6906/Activites-pedagogie-active-IA.pdf
54. Wang, H., Dang, A., Wu, Z., Mac, S.: Generative AI in higher education: Seeing ChatGPT through universities' policies, resources, and guidelines. Computers and Education: Artificial Intelligence **7**, 100326 (Dec 2024). https://doi.org/10.1016/j.caeai.2024.100326, https://www.sciencedirect.com/science/article/pii/S2666920X24001292
55. Yan, L., Greiff, S., Teuber, Z., Gašević, D.: Promises and challenges of generative artificial intelligence for human learning. Nature Human Behaviour **8**(10), 1839–1850 (Oct 2024). https://doi.org/10.1038/s41562-024-02004-5, https://www.nature.com/articles/s41562-024-02004-5, publisher: Nature Publishing Group
56. Yang, S.J.H., Ogata, H., Matsui, T., Chen, N.S.: Human-centered artificial intelligence in education: Seeing the invisible through the visible. Computers and Education: Artificial Intelligence **2**, 100008 (Jan 2021). https://doi.org/10.1016/j.caeai.2021.100008, https://www.sciencedirect.com/science/article/pii/S2666920X21000023

EIAH
2025 Lille
Environnements Informatiques pour l'Apprentissage Humain

Articles courts

Tirer parti des hallucinations en réévaluant la stratégie de l'agent pédagogique perturbateur à l'ère de l'IA générative

Emmanuel G. Blanchard[1,3] [0009-0006-0454-5739], Jean-Christophe Callahan[2] [0000-0002-3376-3982], Idir Akretche[1], Naif Asswiel[1], Lucas Schmitt[1], Amine Kessemtini[1]

[1]Laboratoire d'Informatique de l'Université du Mans, Le Mans, France
[2]Centre Hospitalier du Mans, Le Mans, France
[3]McGill Department of Surgery, Montreal, Canada

emmanuel.blanchard@univ-lemans.fr
jccallahan@ch-lemans.fr

Résumé. L'IA générative en EIAH offre de grandes opportunités, mais elle fait également face des défis majeurs, notamment les risques liés aux hallucinations inhérentes à cette technologie. Dans cet article, nous explorons comment l'IA générative peut être utilisée pour revisiter la stratégie de l'agent perturbateur, une démarche techno-pédagogique inspirée de la théorie de la dissonance cognitive de Festinger. Notre approche tire parti des hallucinations en les transformant en éléments pédagogiques. Elle apparaît particulièrement prometteuse dans le domaine de la formation médicale, où les besoins éducatifs sont importants.

Mots-clés : IA générative, modèle de langage, agent perturbateur, dissonance cognitive, formation médicale.

Abstract. The use of generative AI in EIAH offers significant opportunities but also faces major challenges, particularly the risks associated with the inherent hallucinations of this technology. In this article, we explore how generative AI can be used to revisit the strategy of the Troublemaker agent, a techno-pedagogical approach inspired by Festinger's theory of cognitive dissonance. Our approach leverages hallucinations by transforming them into pedagogical elements. It appears particularly promising in the field of medical education, where educational needs are important.

Keywords: Generative AI, language model, troublemaker agent, cognitive dissonance, medical training.

1 Introduction

Un système de santé efficace repose sur des professionnels bien formés et en nombre suffisant. Or, dans de nombreux pays, trop peu de praticiens ont été formés ces dernières décennies pour répondre au renouvellement nécessaire et à la demande croissante. En France, la loi santé de 2019 a supprimé le numerus clausus [21], mais l'afflux d'étudiants se heurte à un manque d'experts disponibles pour les former, ces derniers étant aussi mobilisés pour les soins.

Les technologies éducatives peuvent aider en soutenant les formateurs [5]. L'un des axes de recherche les plus actifs actuellement est celui visant à identifier des opportunités éducatives liées à l'IA générative. Cependant, il présente des limites, notamment le risque d'hallucinations, i.e. des informations incorrectes pouvant être présentées comme vraies. Cet article propose de transformer cette faiblesse en atout en revisitant la stratégie techno-pédagogique de l'agent perturbateur [1], issue de la théorie de la dissonance cognitive [9]. L'idée est de confronter l'apprenant à des affirmations contradictoires ou déroutantes afin qu'il mobilise ses capacités cognitives pour confirmer ou d'infirmer ses connaissances. Bien encadrée et facilitée par l'IA, cette approche pourrait être particulièrement efficace en formation médicale.

Dans la section 2, nous présentons un court état de l'art de l'apprentissage par perturbation et de l'emploi de l'IA générative en éducation. La section 3 discute des résultats d'une étude que nous avons menée auprès d'experts médicaux afin d'évaluer la plus-value potentielle de l'IA générative dans la mise en place d'activités d'apprentissage par perturbation en contexte médical. Ces résultats et échanges ont influencé le développement de la *Troublemaker Agent Platform* (TrAP), un Environnement Informatique pour l'Apprentissage Humain (EIAH) qui facilite la mise en place et la gestion d'activités d'apprentissage par perturbation. Celui-ci est présenté brièvement dans la section 4. Enfin, en conclusion, nous abordons les travaux futurs que nous anticipons.

2 État de l'Art

Apprentissage par perturbation. Festinger définit la *dissonance cognitive* comme un inconfort mental survenant lorsqu'un individu est confronté à des contradictions entre ses croyances, attitudes ou comportements [9]. Dans cet article, nous regroupons sous l'appellation *apprentissage par perturbation* toutes les approches pédagogiques basées sur la théorie de la dissonance cognitive pour transmettre des connaissances ou des compétences. Celle-ci a été explorée dans plusieurs contextes éducatifs, tels que l'apprentissage des langues [19], la chimie [8], la sensibilisation à la diversité [18] ou encore la promotion du e-learning auprès des enseignants [7]. À la fin des années 1990, des chercheurs ont introduit le concept *d'agent pédagogique perturbateur* [1], un assistant virtuel conçu pour générer volontairement des conflits cognitifs afin de stimuler la pensée critique des apprenants. Bien que prometteuse, le déploiement de cette technique est demeuré limité du fait de limitations technologiques de l'époque.

IA générative et apprentissage. L'usage de grands modèles de langage (LLM, Large Language Model) en éducation est récent mais attise un fort intérêt de la communauté de recherche en EIAH [10]. Des travaux ont ainsi mis en évidence que des chatbots basés sur des LLM peuvent contribuer efficacement à des activités de tutorat et à l'évaluation de réponses courtes [20], agir en tant que partenaires virtuels [17], faciliter la génération de questions pédagogiques à grande échelle [6] ou que l'IA générative peut contribuer à la conception de curriculums pédagogiques [22]. Toutefois, l'intégration de l'IA générative dans les environnements d'apprentissage soulève plusieurs défis [13], notamment en matière de fiabilité, de transparence [24] et de biais culturels [4]. L'un des problèmes majeurs est celui des « hallucinations », c'est-à-dire la production d'informations erronées présentées comme véridiques [11, 12]. La génération augmentée par récupération (Retrieval-Augmented Generation – RAG) constitue une solution largement adoptée pour limiter ce phénomène [14]. Elle consiste à enrichir les prompts de génération à l'aide de contenus provenant d'une base de connaissances. Cela permet d'avoir un meilleur contrôle sur la génération de texte [15, 16, 23] et peut ainsi réduire les risques liés aux hallucinations. De ce fait, l'approche RAG connaît un intérêt croissant dans le domaine des technologies éducatives intelligentes.

Chacune de ces approches présente des opportunités pour mieux prendre en compte les faiblesses de l'autre, ce pourquoi nous pensons que leur association permettrait des technologies éducatives innovantes et efficaces.

3 Évaluation

L'équipe à l'origine de cette recherche est interdisciplinaire et se compose d'un médecin réanimateur et d'un chercheur spécialisé en intelligence artificielle pour l'éducation (AIED). La principale question de recherche de notre évaluation est la suivante :

> **Un chatbot à base de LLM non spécialisé peut-il générer des affirmations médicales fausses mais suffisamment réalistes pour être utilisables dans des activités d'apprentissage par perturbation ?**

Cette notion de suffisamment réaliste est critique pour espérer une plus-value pédagogique car une affirmation trop évidemment fausse ne provoquera pas le doute nécessaire à l'état de dissonance cognitive au cœur de l'apprentissage par perturbation. De même, si toutes les affirmations proposées dans une activité d'apprentissage sont fausses, l'étudiant pourrait identifier ce point et contourner les objectifs pédagogiques (i.e. « *gaming the system* » [2]). Dans une activité d'apprentissage par perturbation, il est donc nécessaire que les affirmations fausses soient suffisamment réalistes et qu'elles soient mélangées avec des affirmations vraies afin d'induire un doute suffisant chez l'étudiant au moment de l'activité pédagogique.

Nous avons commencé par effectuer des tests préliminaires au moyen de la version en accès libre de ChatGPT (modèle ChatGPT-4o, période de novembre/décembre 2024). Notre prompt contenait une simple question sans exemple de réponse (technique dite de *zero-shot prompting*) en vue de générer 3 affirmations distinctes fausses mais

réalistes. Nous avons réalisé quelques essais afin de raffiner ce prompt, l'expert médical jugeant le côté faux mais réaliste des réponses médicales générées. Le prompt finalement retenu est le suivant :

> *« En reprenant la question que je vais te fournir par la suite, génère 3 affirmations a) distinctes et b) fausses mais réalistes pour un médecin réanimateur expert.»*

Ainsi pour la question médicale « *Quels sont les moyens de prévention de la pneumopathie acquise sous ventilation mécanique qui ont fait la preuve de leur efficacité quand utilisés en première intention ?* », le chatbot a généré trois affirmations parmi lesquelles la suivante :

> *« L'administration systématique de corticostéroïdes pour réduire l'inflammation pulmonaire est une méthode efficace pour prévenir la pneumopathie sous ventilation mécanique.»*

Comme escompté, cette réponse est médicalement fausse mais l'explication est effectivement vraie. Ces tests préliminaires ont été jugés suffisamment prometteurs pour que nous mettions en place une évaluation en vue d'obtenir un retour plus large de la communauté médicale.

Méthode. L'expert médical de notre équipe a identifié 11 questions médicales, à partir desquelles 33 affirmations ont été générées selon l'invite précédente.

Un panel d'experts médicaux a ensuite évalué le ***degré de fausseté*** de ces affirmations en les notant à l'aide de l'échelle suivante : *toujours faux, généralement faux, généralement vrai, toujours vrai*. Les experts pouvaient également choisir de ne pas répondre si l'énoncé sortait de leur champ d'expertise.

Par la suite, ils ont évalué le ***potentiel de perturbation*** de chaque affirmation, c'est-à-dire la probabilité qu'elle suscite un doute chez deux types différents d'utilisateurs — les externes en médecine et les internes[1] — à l'aide d'une échelle de Likert en cinq points (*jamais, rarement, parfois, souvent, toujours*).

Le panel a enfin répondu à des questions ouvertes, afin d'analyser plus en profondeur la qualité des affirmations générées ainsi que la pertinence de l'apprentissage par perturbation dans la formation médicale.

Description du panel d'experts. L'évaluation s'est déroulée sur une période de deux semaines, entre la fin de l'année 2024 et le début de l'année 2025, et a mobilisé quatorze experts médicaux (douze hommes et deux femmes) exerçant dans un hôpital français. Treize d'entre eux travaillent en soins intensifs et un est cardiologue. Leur expérience moyenne en tant que médecins seniors est de 14,4 ans (écart-type : 12,7). Tous les participants encadrent régulièrement des externes et des internes en médecine. Le nombre

[1] Les externes sont des étudiants dans les premières années de la formation en médecine alors que les internes (parfois appelés résidents) sont de jeunes médecins encore encadrés par d'autres professionnels plus experts.

important d'experts ayant accepté de participer à notre étude, malgré des emplois du temps très contraints, témoigne de l'intérêt marqué suscité par notre initiative.

Résultats. Les experts ont d'abord évalué la capacité de ChatGPT-4o à générer des affirmations fausses. Tous (100 %) se sont accordés à reconnaître que 20 des 33 affirmations étaient fausses, tandis que 75 à 99 % d'entre eux ont confirmé la fausseté de 11 autres affirmations. Les deux dernières affirmations ont fait l'objet de jugements plus nuancés, avec entre 50 et 74 % les identifiant comme fausses.

Le potentiel de perturbation de ces affirmations a ensuite été examiné. Les réponses du panel ont été classées selon quatre catégories : très probable, probable, peu probable et très peu probable de susciter de la confusion. Les affirmations ne présentant pas de tendance claire n'ont pas été catégorisées.

Le potentiel de perturbation a été jugé plus élevé chez les externes (1 très probable, 8 probables, 4 peu probables, 0 très peu probable) que chez les internes (0 très probable, 3 probables, 12 peu probables, 1 très peu probable). Cette différence de jugement du panel était attendue étant donné que les externes sont beaucoup plus novices que les internes. Si ces performances sont pour le moment modestes (27% pour les externes et 9% pour les internes), ces résultats permettent de valider H1 : un chatbot à base de LLM non spécialisé peut générer des affirmations fausses mais suffisamment réalistes pour être utilisé dans de l'apprentissage médical par perturbation. De plus, des commentaires additionnels de nos experts mettent en avant que des interventions limitées sur les affirmations générées (révisions grammaticales, ajustement de quelques termes) amélioreraient grandement le potentiel de perturbation de nombreuses affirmations. Dans la même idée, l'expert médical de notre équipe a remarqué qu'en demandant au même LLM d'améliorer la première affirmation générée, il obtenait des résultats au potentiel de perturbation bien plus évident. Ces éléments additionnels pointent l'intérêt de garder l'humain dans la boucle afin de tirer le meilleur parti d'un support créatif à base d'IA générative.

Nous avons également analysé la cohérence entre les jugements des évaluateurs en calculant l'α de Krippendorf [14]. En plus d'être adaptée à des jugements obtenus à partir d'une échelle ordinale (i.e. Likert), cette métrique présente plusieurs avantages pour notre cas spécifique comme la capacité à être calculée sur de nombreux juges (14 dans notre cas) et à prendre en compte des données manquantes (nous avons codé comme manquante les situations où un expert jugeait l'affirmation hors de son champ d'expertise ou trop avancée pour la classe d'étudiants visée). Le tableau 1 présente notre évaluation de la concordance de jugement des experts médicaux.

Table 1. Évaluation de la concordance de jugement des évaluateurs

Potentiel de perturbation	α de Krippendorf	IC 95% (1000 itérations)
Fausseté des affirmations	0.151	[0.067, 0.231]
Potentiel de perturbation pour les externes	0.179	[0.063, 0.294]
Potentiel de perturbation pour les internes	0.240	[0.123, 0.379]

Les valeurs obtenues sont relativement faibles avec des α compris entre 0.151 et 0.240, indiquant un accord limité entre les évaluateurs. Ces résultats suggèrent que l'accord observé est supérieur à celui attendu par le hasard sans pour autant garantir une forte fiabilité inter-juges.

Selon l'expert médical de l'équipe, une part de ces divergences pourraient être dues à un manque d'attention de certains répondants. En effet, la véracité de plusieurs affirmations dépendait de quelques termes précis, qui n'ont peut-être pas été bien pris en compte. Par ailleurs, les affirmations portaient uniquement sur la réanimation. Bien qu'il s'agisse de notions de base, certains experts (e.g. cardiologie) n'y sont plus confrontés dans leur pratique actuelle. En outre, certaines affirmations reflétaient des recommandations pouvant varier selon les sociétés médicales. Dépendamment de son niveau d'informations quant à ces divergences, un expert pouvait fournir des évaluations différentes mais correctes en fonction de l'interprétation. Par ailleurs, l'évaluation du potentiel de perturbation des affirmations reste aussi subjective, car elle dépend du ressenti de chaque expert quant à l'effet sur les étudiants. Il sera donc important de comparer ce ressenti avec les résultats réels observés chez les étudiants.

En résumé et en suivant l'analyse de l'expert médical de notre équipe, nous pensons que les faibles valeurs des α de Krippendorf mettent avant tout en avant la dimension personnelle de l'approche d'enseignement de chacun des 14 experts médicaux de notre panel. Cela confirme également l'intérêt de garder les experts humains dans la boucle, afin que chacun puisse obtenir des générations alignées sur son style d'enseignement.

4 L'environnement TrAP (Troublemaker Agent Platform)

Les résultats de notre évaluation précédente ont confirmé l'intérêt de développer la Troublemaker Agent Platform (TrAP). Cette section propose une courte présentation de cet environnement informatique pour favoriser la mise en place d'activités d'apprentissage par perturbation en utilisant l'IAGen comme support créatif tout en gardant l'expert dans la boucle. Des réflexions complémentaires nous ont conduits à définir les objectifs suivants pour ce système :

- Réduire le temps et les efforts nécessaires à un expert pour concevoir, gérer et analyser des activités d'apprentissage par perturbation, tout en lui garantissant la possibilité de contrôler et de modifier ces activités selon ses besoins ;
- Développer un module, fondé sur l'IAGen, capable de produire des affirmations de qualité, tout en étant peu coûteux, respectueux de l'environnement et garantissant la protection des données.

L'image 1 présente une architecture modulaire simplifiée de la plateforme TrAP.

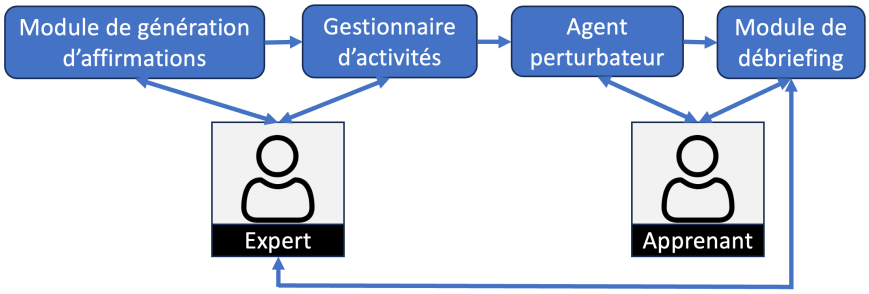


Image 1. Une architecture modulaire simplifiée de TrAP

Le module de génération d'affirmations assiste les experts dans la création d'énoncés pertinents pour l'apprentissage par perturbation. Le gestionnaire d'activités permet à l'expert de modifier les affirmations générées, d'en créer de nouvelles sans recourir à l'intelligence artificielle, et de conserver celles qui lui conviennent in fine. Il est également utilisé par l'expert pour regrouper plusieurs énoncés en une activité d'apprentissage. L'agent perturbateur présente ensuite cette activité à l'apprenant via une interface simple (image 2), laquelle affiche une affirmation médicale (A), demande à l'apprenant de juger de sa véracité (B), puis de justifier son raisonnement (C).

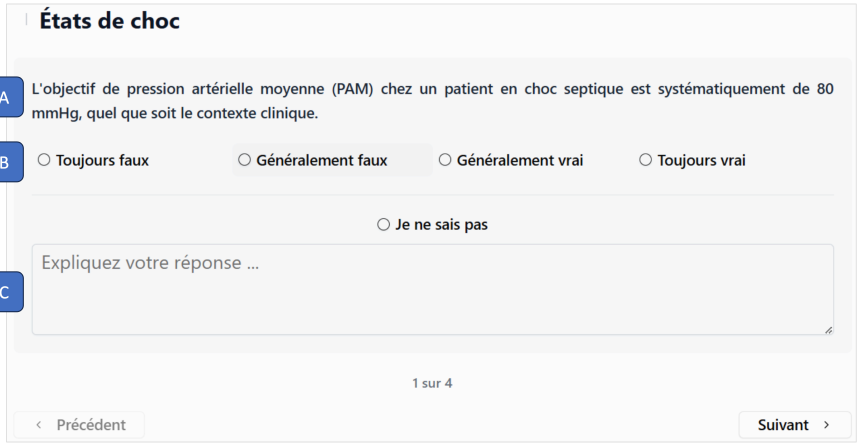


Image 2. L'interface graphique de l'agent perturbateur de TrAP

Les traces d'interaction et les réponses de l'utilisateur sont ensuite exploitées par le module de débriefing afin de soutenir une séance de rétroaction entre l'expert et l'apprenant.

5 Conclusion et travaux futurs

Dans cet article, nous avons validé que l'IAGen peut contribuer à la mise en œuvre d'activités d'apprentissage par perturbation en apportant un soutien créatif aux experts, dans un contexte médical. L'apprentissage par perturbation permet un usage constructif d'affirmations fausses mais plausibles. Par conséquent, les hallucinations inhérentes à l'IAGen ne constituent plus un problème dès lors qu'elles sont encadrées de manière appropriée, c'est-à-dire en laissant le contrôle final à l'expert. Faciliter la mise en place de telles activités est l'objectif de l'environnement TrAP que nous avons succinctement présenté. Cet environnement entend répondre à plusieurs défis associés à l'apprentissage par perturbation et à l'usage de l'IAGen.

Ainsi, afin de garantir la soutenabilité financière et écologique de cette, nous développons actuellement des variantes du module de génération d'affirmations reposant sur des petits langages de modèle et un processus RAG. Des évolutions préliminaires de cette démarche ont montré qu'il est possible, dans le contexte de l'apprentissage par perturbation, d'atteindre des performances proches de celles d'un modèle de langage commercial avec ce type d'architecture, ce qui offre une solution plus économique et assurant la souveraineté des données [3].

Dans le cadre de l'apprentissage par perturbation, les séances de débriefing sont essentielles pour prévenir l'internalisation de fausses informations. Bien que la systématicité du débriefing soit commune l'enseignement médical — et ne suscite donc pas d'inquiétude particulière selon notre expert médical —, cette approche est moins fréquente dans d'autres domaines et il apparait important d'insister sur ce point.

Pour faciliter les sessions de débriefing entre experts et apprenants, TrAP intègre un module. Nos travaux futurs porteront notamment sur l'enrichissement de ces séances par des visualisations de données. L'évaluation à grande échelle de la plateforme TrAP, ainsi que la conception d'un module de génération d'affirmations vraies mais troublantes sont également des objectifs à moyen terme.

References

1. Aimeur, E., Frasson, C., Dufort, H.: Cooperative learning strategies for intelligent tutoring systems. Applied Artificial Intelligence. 14, 465–489 (2000). https://doi.org/10.1080/088395100403388.
2. Baker, R.S.J.D., Corbett, A.T., Koedinger, K.R., Evenson, S., Roll, I., Wagner, A.Z., Naim, M., Raspat, J., Baker, D.J., Beck, J.E.: Adapting to when students game an intelligent tutoring system. In: Procs on ITS2006, Springer LNCS 4053. pp. 392–401 (2006). https://doi.org/10.1007/11774303_39.
3. Blanchard, E.G, Callahan, J-C, Akretche, I., Asswiel, N., Schmitt, L., Kessemtini, A.,Khan, I.S.A.: Making Generative AI Hallucinations Useful by Reassessing the Troublemaker Agent Strategy. To appear in Late Breaking Results of the 26th International Conference on Artificial Intelligence in Education (2025).
4. Blanchard, E.G., Mohammed, P.: On Cultural Intelligence in LLM-Based Chatbots: Implications for Artificial Intelligence in Education. In: Procs of AIED2024, Springer LNCS 14829. pp. 439–453 (2024). https://doi.org/10.1007/978-3-031-64302-6_31.

5. Bridges, S., Chan, L.K., Hmelo-Silver, C.E.: Educational technologies in medical and health sciences education. (2015). https://doi.org/10.1007/978-3-319-08275-2.
6. Bulathwela, S., Muse, H., Yılmaz, E.: Scalable Educational Question Generation with Pre-trained Language Models. In: Artificial Intelligence in Education. pp 327–339 (2023).
7. Chiou, W.B.: Using Cognitive Dissonance to Enhance Faculty Members' Attitudes toward Teaching Online Courses. Psychological Reports, 99 (2), 465-471 (2006).
8. Corradi, D., Clarebout, G. & Elen, J.: Cognitive Dissonance as an Instructional Tool for Understanding Chemical Representations. J Sci Educ Technol 24, 684–695 (2015). https://doi.org/10.1007/s10956-015-9557-5
9. Festinger, L.: The arousal and reduction of dissonance in social contexts. In: S. Schachter and M. Gazzaniga (eds), Extending Psychological Frontiers: Selected Works on Leon Festinger. pp. New York : Russell Sage Foundations. pp 238-257 (1989).
10. Gilliot J. M. & Pentecouteau, H.: Table ronde : IA et éducation à l'heure de ChatGPT. In: Procs of EIAH2023 : 11ème Conférence sur les Environnements Informatiques pour l'Apprentissage Humain. pp 18-19 (2023).
11. Hicks, M.T., Humphries, J., Slater, J.: ChatGPT is bullshit. Ethics and Information Technology. 26, (2024). https://doi.org/10.1007/s10676-024-09775-5.
12. Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., ... & Fung, P.: Survey of hallucination in natural language generation. ACM Computing Surveys, *55*(12), 1-38 (2023).
13. Kooli, C.: Chatbots in education and research: A critical examination of ethical implications and solutions. *Sustainability*, *15*(7), p. 5614 (2023).
14. Lewis, P.S.H., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.-T., Rocktäschel, T., Riedel, S., Kiela, D.: Retrieval-Augmented Generation for Knowledge-Intensive NLP tasks. Neural Information Processing Systems. 33, 9459–9474 (2020).
15. Li, X., Henriksson, A., Duneld, M., Nouri, J., Wu, Y.: Supporting Teaching-to-the-Curriculum by Linking Diagnostic Tests to Curriculum Goals: Using Textbook Content as Context for Retrieval-Augmented Generation with Large Language Models. In: Olney, A.M., Chounta, IA., Liu, Z., Santos, O.C., Bittencourt, I.I. (eds) Artificial Intelligence in Education. AIED 2024., vol 14829. Springer, Cham. (2024) https://doi.org/10.1007/978-3-031-64302-6_9
16. Liu, C., Hoang, L., Stolman, A., Wu, B.: HiTA: A RAG-Based Educational Platform that Centers Educators in the Instructional Loop. In: Olney, A.M., Chounta, IA., Liu, Z., Santos, O.C., Bittencourt, I.I. (eds) Artificial Intelligence in Education. AIED 2024., vol 14830. Springer, Cham. (2024) https://doi.org/10.1007/978-3-031-64299-9_37
17. Ma, Q. C., Wu, S. T., Koedinger, K: Is AI the better programming partner? Human-human pair programming vs human-AI pAIr programming. In Proc. of the AIED Workshop on Empowering Education with LLMs - the Next-Gen Interface and Content Generation, pp 64-77 (2023).
18. McFalls, E. L., & Cobb-Roberts, D.: Reducing Resistance to Diversity through Cognitive Dissonance Instruction: Implications for Teacher Education. Journal of Teacher Education, 52(2), 164-172 (2001). https://doi.org/10.1177/0022487101052002007
19. Miao, Y., Moran, M., & Kang, O.: A cognitive dissonance approach to moderating listener perception of L2 English speakers. Journal of Multilingual and Multicultural Development, pp. 1–18 (2023). https://doi.org/10.1080/01434632.2023.2236073
20. Nye, B. D., Mee, D., Core, M.G.: Generative Large Language Models for Dialog-Based Tutoring: An Early Consideration of Opportunities and Concerns. In Proc. of the AIED Workshop on Empowering Education with LLMs - the Next-Gen Interface and Content Generation, pp 78-88 (2023).

21. Roberto, A. : La fin du numérus clausus n'aura pas d'effet avant 2035. Fondation pour la recherche sur les administrations et les politiques publiques. https://www.ifrap.org/emploi-et-politiques-sociales/la-fin-du-numerus-clausus-naura-pas-deffets-avant-2035. Dernier accès le 13 janvier 2025.
22. Sridhar, P. Doyle, A., Agarwal, A. Bogart, C., Savelka, J., Sakr, M. Harnessing LLMs in curricular design. Using GPT-4 to support authoring of learning objectives. In Proc. of the AIED Workshop on Empowering Education with LLMs - the Next-Gen Interface and Content Generation, pp 139-150 (2023).
23. Thüs, D., Malone, S., Brünken, R.: Exploring generative AI in higher education: a RAG system to enhance student engagement with scientific literature. Front. Psychol. 15, 1474892 (2024). https://doi.org/10.3389/fpsyg.2024.1474892.
24. Wu, Y., Henriksson, A., Duneld, M., Nouri, J.: Toward improving the reliability and transparency of ChatGPT for educational question answering. In Proc. of EC-TEL2023, Springer LNCS 14200, pp 475-488 (2023).

Etude des effets d'un dispositif de formation hybride sur les compétences à s'orienter des étudiants du supérieur

Madina, Bouamama[1], Rawad, Chaker[1], Stéphane, Simonian[1]

[1] ECP, Université Lumière Lyon 2
madina.bouamama@univ-lyon2.fr
rawad.chaker@univ-lyon2.fr
stephane.simonian@univ-lyon2.fr

Résumé. Cette étude porte sur l'évaluation des compétences à s'orienter des étudiants du supérieur dans un dispositif de formation hybride intégré au cursus de licence, dans le cadre d'un projet de recherche national axé sur la maturité vocationnelle. Une enquête menée auprès de 455 étudiants en pré-test et de 186 en post-test a permis de concevoir un outil d'évaluation et d'analyser la progression des étudiants. Parmi ces derniers, 117 ont participé à la fois aux pré-tests et aux post-tests, sur lesquels nous avons réalisé une analyse approfondie de l'impact du dispositif de formation sur leur apprentissage. Les résultats de cette expérimentation seront confrontés à des éléments théoriques concernant les effets des dispositifs hybrides sur les apprentissages.

Mots-clés : compétences à s'orienter, dispositif de formation, outil d'évaluation, hybridation, enseignement supérieur.

Abstract. This study focuses on the assessment of the career guidance skills of higher education students in a hybrid training system integrated into the bachelor's degree curriculum, as part of a national research project focused on vocational maturity. A survey conducted among 455 students in pre-test and 186 in post-test made it possible to design an assessment tool and analyze the students' progress. Among the latter, 117 participated in both the pre-tests and post-tests, on which we carried out an in-depth analysis of the impact of the training system on their learning. The results of this experiment will be compared with theoretical elements concerning the impact of hybrid systems on educational stakeholders.

Keywords: career management skills, training system, assessment tool, training system, higher education

1 Introduction

Notre étude porte sur l'évaluation d'un dispositif hybride dans l'enseignement supérieur visant à développer les compétences à s'orienter d'étudiants de première et de deuxième année. Nous mobilisons un outil d'évaluation (auteurs, en cours d'évaluation)

avec pour objectif de mesurer la capacité des étudiants à développer l'autonomie indispensable à l'acquisition des compétences d'orientation et à la préparation de leur avenir professionnel.

Dans ce cadre, nous avons retenu la définition dispositionnelle de la compétence à s'orienter, telle que proposée par Michaud (2003), qui conçoit la compétence à s'orienter comme un « savoir-agir complexe ». Ce savoir-agir se traduit par la mobilisation des ressources personnelles, professionnelles et environnementales nécessaires pour relever les défis liés à l'orientation. À l'instar de Le Boterf (2000), Michaud met en avant que cette compétence ne se résume pas à l'accumulation de ressources, mais qu'elle réside dans la capacité à les mobiliser de manière efficace (Michaud, 2003). Dans ce contexte, nous avons mené une étude longitudinale portant sur la progression en termes de compétences à s'orienter entre avant et après la formation, et ainsi évaluer les effets du dispositif hybride sur l'appropriation de ces compétences. Plusieurs études montrent que la manière de percevoir les qualités instrumentales d'un dispositif de formation à distance, peut influencer l'engagement et l'apprentissage des étudiants (Caron et al. 2017 ; Molinari et al., 2016).

Partant, notre question de recherche est la suivante : dans quelle mesure la perception des qualités instrumentales du dispositif hybride a influencé l'acquisition de compétences à s'orienter ?

Comprendre les interactions entre facteurs d'acceptabilité du dispositif et la montée en compétences dans un domaine, permettrait non seulement de mieux accompagner les étudiants dans cette transition, mais aussi de réfléchir aux pratiques pédagogiques à adopter pour favoriser une meilleure intégration des outils et méthodes hybrides dans l'enseignement supérieur. Le présent article rend compte de l'utilisation d'un outil d'évaluation spécifiquement conçu pour ce dispositif.

2 Contexte de l'étude

Le dispositif de formation SCALE pour l'acquisition des compétences à s'orienter s'inscrit dans le cadre d'un projet national financé par le programme DEFFINUM (2021-2025), dédié à l'innovation numérique au service de la formation. Cette initiative bénéficie du soutien de l'État dans le cadre de l'Appel à Manifestation d'Intérêt (AMI) « Compétences et Métiers d'Avenir » du Programme France 2030, opéré par la Caisse des Dépôts. Expérimenté au premier semestre 2023 auprès de 3000 étudiants de l'Université Claude Bernard Lyon 1, le dispositif est actuellement en phase de généralisation à l'échelle nationale. En s'appuyant sur la définition de l'hybridation proposée par Peraya et al. (2014) qui souligne l'articulation entre des activités en présentiel et à distance pour créer une synergie pédagogique, ce dispositif intègre des modalités hybrides dans sa conception. Il vise à renforcer les compétences des étudiants en matière d'orientation et s'intègre aux modules de licence PPP (Parcours Personnel et Professionnel), organisés en deux volets : PPP2 et PPP4. Les contenus pédagogiques de ces modules sont co-développés par l'équipe pédagogique du Service d'Orientation et d'Insertion professionnelle des Étudiants (SOIE) de l'Université Claude Bernard Lyon 1 et les experts de l'entreprise JobTeaser. L'évaluation des modules se caractérise par une présentation

orale dans un format synthétique. Le module PPP2 introduit une démarche active de réflexion sur l'avenir professionnel des étudiants, en les engageant dans l'exploration d'un métier ou d'un secteur professionnel. Ils sont amenés à réaliser des recherches documentaires et à mener des entretiens avec des professionnels. Par exemple, le contenu du cours « développer un réseau ça s'apprend » leur permet de développer des compétences en recherche et en traitement de l'information par le biais d'interactions avec des acteurs du monde professionnel. Le PPP4 les accompagne dans l'élaboration et la présentation de leur projet professionnel et inclut l'articulation entre le parcours académique et professionnel, permettant aux étudiants d'apprendre à structurer leur présentation et à défendre leur projet. Les contenus comprennent des modules en ligne tels que « Se raconter », « Identifier ses compétences », suivis des TD. La figure suivante illustre la scénarisation de ce module, spécialement conçu pour accompagner les étudiants dans l'identification et la valorisation de leurs compétences.

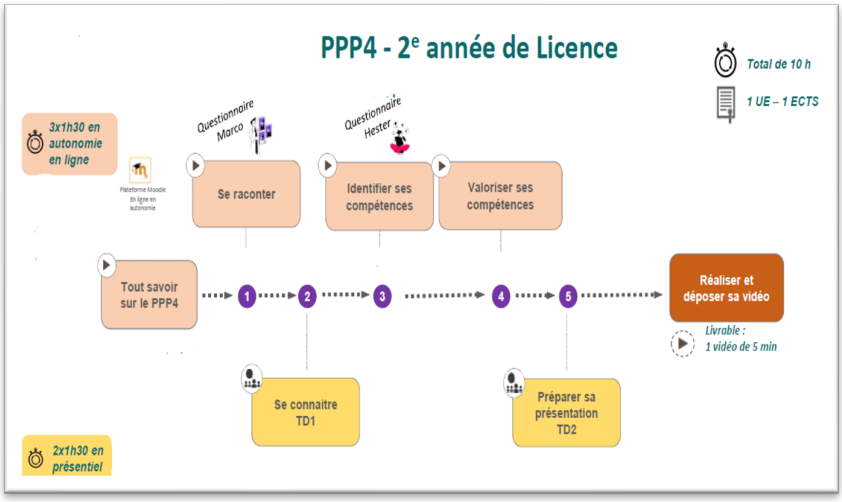


Fig. 1. Scénarisation pédagogique du module PPP4 dans Moodle

3 Evaluation des effets du dispositif de formation hybride sur l'apprentissage des étudiants

Pour mesurer les effets du dispositif de formation sur l'apprentissage des étudiants, les questionnaires pré- et post-formation explorent des aspects spécifiques du parcours d'orientation, allant de l'acquisition des savoir-faire nécessaires à la gestion de leur parcours d'orientation, jusqu'à l'évaluation de la perception du dispositif numérique mis à leur disposition. Les questions sont mesurées à l'aide d'une échelle de Likert à 7 points, allant de « pas du tout d'accord » à « tout à fait d'accord ».

3.1 Questionnaire pré-formation

Le questionnaire utilisé a été validé dans le cadre d'une étude préliminaire après des analyses factorielles exploratoires et confirmatoires (auteurs, en cours d'évaluation) et est structuré autour de trois dimensions principales :

Les savoir-faire
Cette dimension se compose de 20 items et évalue l'identification des compétences nécessaires pour développer des savoir-faire clés en orientation, telles que définies par l'outil d'évaluation professionnel SavOrient (Beltrame, 2019). Cet outil s'aligne avec le référentiel des huit compétences de Pôle emploi (Pôle emploi, 2013), qui couvre trois grandes thématiques : la connaissance de soi, la connaissance du marché de l'emploi et des entreprises, ainsi que la connaissance du système de formation et des certifications. Les sous-dimensions associées aux compétences du SavOrient sont les suivantes : Se découvrir et identifier ses compétences (SDIC), Identifier des soutiens dans son environnement pour son projet (IESP), Expliciter son expérience, ses choix et son projet (EECP), Élaborer et mettre en œuvre son plan d'action (EMPA), et Activer et élargir son réseau (AESR).

Les préoccupations
Cette dimension porte sur les préoccupations professionnelles des étudiants, évaluées à travers 29 items. Elle s'appuie sur le modèle de Super (1955), qui est inspiré de la théorie des tâches développementales de Havighurst (1953), afin de mesurer l'évolution des préoccupations des étudiants en matière d'orientation professionnelle. Ce modèle a été redéfini dans le cadre de notre étude préliminaire en quatre étapes, chacune associée à des objectifs spécifiques. La première phase, « IO » (Identifier des Opportunités), concerne la découverte des options de carrière et l'exploration des possibles. La deuxième phase, « ACA » (Améliorer ses Chances d'Avancement), se concentre sur la planification et la mise en œuvre de stratégies pour progresser dans un domaine choisi. La troisième phase, « SPAC » (Se Perfectionner pour s'Adapter au Changement), met l'accent sur l'importance de développer de nouvelles compétences pour s'adapter aux évolutions professionnelles. Enfin, la quatrième phase, « EOA » (Exploration des Options Alternatives), incite les étudiants à réfléchir à des alternatives professionnelles ou à des changements de parcours.

Le Sentiment d'Efficacité Personnelle lié à l'Expérience
Le Sentiment d'Efficacité Personnelle lié à l'Expérience (SEPE) mesure l'impact des expériences vécues par les étudiants sur leur confiance en leur capacité à atteindre leurs objectifs. Pour ce faire, nous avons adapté l'échelle d'Efficacité Personnelle en Formation de Heutte (2011), fondée sur la théorie sociale cognitive de Bandura (1997), qui souligne que les attentes en matière d'efficience personnelle sont influencées par les expériences passées et peuvent modifier les comportements. Cette dimension s'aligne avec notre approche dispositionnaliste, et met en lumière l'importance de la réflexion sur ses expériences. Il s'agit d'un processus essentiel pour le développement des compétences d'auto-évaluation, la capacité à interpréter ses expériences étant un facteur clé pour la gestion de carrière (Heath, 1960).

3.2 Questionnaire post-formation

Le questionnaire post-formation reprend les trois dimensions abordées lors du pré-test dans une visée longitudinale, et intègre en plus quatre dimensions visent à évaluer la manière dont le dispositif de formation hybride a été perçu. Cette dimension évalue la manière dont les apprenants perçoivent le dispositif de formation hybride. Elle mobilise 12 items regroupés en quatre principales variables, chacune étant évaluée à l'aide de trois items spécifiques.

Perception du dispositif hybride

La qualité des informations transmises dans le cours (INFO)
Cette dimension met l'accent sur la clarté et la pertinence des informations fournies dans le cours, permettant aux apprenants de mieux s'approprier les contenus pédagogiques (Poellhuber & Roy, 2018)

L'engagement des apprenants dans les activités proposées (ENG)
Cette composante s'intéresse à l'interactivité et à la participation active des apprenants dans les activités proposées au sein du cours, favorisant une meilleure implication dans l'apprentissage (Poellhuber & Roy, 2018)

L'utilité perçue des outils numériques (UTIL)
Cette variable évalue la perception globale de l'utilité des outils numériques pour soutenir l'apprentissage, ainsi que leur pertinence dans le cadre des formations à distance (Caron & Heutte, 2017)

La dimension hybride (présentiel/distanciel) de la formation (PRES)
Cette dimension se concentre sur la complémentarité entre les modalités de formation en présentiel et à distance, soulignant la capacité d'adaptation des apprenants et des formateurs dans des contextes hybrides. Les items ont été créées par les chercheurs.

4 Méthodologie

Nous décrivons dans cette section le processus de collecte de données ainsi que les caractéristiques de l'échantillon étudié. Ensuite, nous détaillons les méthodes statistiques utilisées pour l'analyse de ces données.

4.1 Échantillon

Le questionnaire, conçu et validé par des experts, a été administré en ligne auprès des étudiants inscrits aux modules de licence 1 et 2, PPP2 et PPP4, avant et après la formation. La collecte des données s'est réalisée via la plateforme LimeSurvey, où les participants ont dû donner leur consentement pour que les données recueillies, anonymisées, soient utilisées exclusivement dans le cadre de cette recherche. Le nombre de participants en pré-formation était de N = 455 et en post-formation de N = 186. Parmi eux, l'échantillon final se compose de N = 117 participants ayant répondu à la fois à l'enquête pré-formation et post-formation, représentant 18,25 % des étudiants concernés

par les deux enquêtes. La répartition des participants montre une légère surreprésentation féminine (57,14 % de femmes) ainsi qu'un équilibre relatif entre les deux modules (53,85 % inscrits en PPP2 et 46,15 % en PPP4). En ce qui concerne les expériences professionnelles, 59,8 % des participants n'ont jamais acquis d'expérience professionnelle. Parmi ceux ayant une expérience, 12,09 % ont occupé un poste en CDI, 10,33 % en CDD, 7,25 % ont effectué un stage, 27,25 % ont mentionné un job d'été, et 3,08 % ont signalé d'autres formes d'expérience. Concernant leur situation actuelle par rapport à l'emploi, 19,34 % des répondants déclarent être employés, 20,22 % recherchent activement un emploi, tandis que 60,22 % ne sont pas en recherche d'emploi.

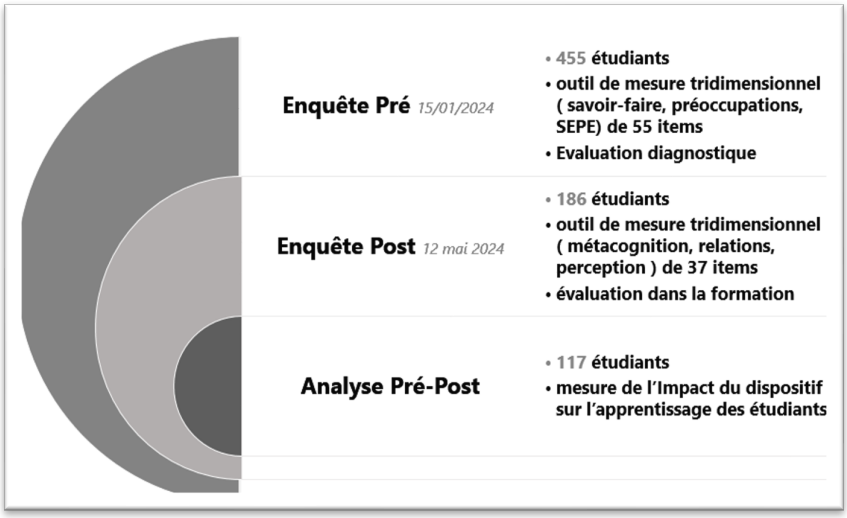


Fig. 2. Étapes de l'enquête menée auprès des étudiants des modules PPP2 et PPP4

4.2 Analyse des données

Avant de procéder aux analyses longitudinales, nous avons procédé à l'alignement des données entre pré- et post-test (échantillon aligné : N = 117). Afin de mesurer le delta d'apprentissage, nous avons d'abord effectué des tests *t* de Student pour échantillons appariés pour chaque dimension et sous-dimension de chaque mesure des compétences à s'orienter. Dans un second temps, nous avons procédé à des ANOVAs à mesures répétées pour chaque dimension, en intégrant les quatre sous-échelles de la perception instrumentale du dispositif hybride, comme co-variable explicative des variations intra-sujets entre pré- et post-test.

5 Résultats

Cette section présente les résultats relatifs à l'impact du dispositif de formation sur l'apprentissage des étudiants, en comparant les données recueillies avant et après la

formation selon trois dimensions : les savoir-faire, les préoccupations et le SEPE. Par la suite, nous analyserons la relation entre l'évolution de ces dimensions avant et après la formation et la perception de l'outil numérique.

5.1 Résultats descriptifs

Les échelles indiquent de bonnes fiabilités internes en pré- et en post-test (voir Tableau 1). Concernant l'acquisition des savoir-faire, une légère augmentation des moyennes a été observée, passant de M = 5,15 ; ET =0,80 avant la formation, à M = 5,26 ; ET = 0,85 après celle-ci. Les résultats détaillés de cette comparaison sont présentés dans le Tableau

Tableau 1 : Statistiques descriptives : comparaison des résultats des tests t entre le pré et post-test

Variables de recherche	**Pré-test**		**Post-test**		***t***	***p***
	M	ET	M	ET		
Savoir-faire	5,15	0,80	5,26	0,85	1,70	0,091
SDIC	5,21	0,89	5,26	0,93	0,59	0,556
IESP	5,37	0,98	5,25	1,02	1,32	0,189
EECP	5,40	1,00	5,52	1,00	1,34	0,180
EMPA	5,56	1,12	4,88	1,12	3,65	0,000
AESR	5,07	1,17	5,37	1,11	3,65	0,107
Préoccupations	4,42	1,01	4,75	0,89	3,70	0,000
IO	4,92	1,18	5,04	1,18	0,99	0,320
ACA	4,94	1,26	5,08	1,16	1,16	0,246
SPAC	4,69	1,31	4,82	1,15	1,05	0,296
EOA	3,13	1,64	4,09	1,22	7,13	0,000
SEPE	4,58	1,00	4,98	0,99	4,70	0,000
Perception instrumentale						
INFO			4,49	1,37		
ENG			4,11	1,53		
UTIL			435	1,65		
PRES			4,42	1,48		

p < 0.001

5.2 ANOVAs à mesures répétées

Cette section présente les résultats relatifs à l'impact du dispositif de formation sur l'apprentissage des étudiants, en comparant les données recueillies avant et après la formation selon trois dimensions : les savoir-faire, les préoccupations et le SEPE.

Les savoir-faire

Pour les savoir-faire, bien qu'une légère amélioration globale a été constatée, cette différence n'est pas significative ($p = 0,091$). Plus en détails, les tests appariés pour chaque sous-dimension ont produit des résultats contrastés. Les résultats indiquent que les sous-dimensions EMPA et EASR progressent significativement (respectivement, $t = 1,70$; $p < 0,01$ et $t =$; $p < 0,01$). Les trois autres sous-dimensions n'indiquent pas de progrès significatif ($p > 0,05$).

Afin d'évaluer l'effet de la perception instrumentale du dispositif, les ANOVAs à mesures répétées indiquent que SDIC interagit significativement avec INFO et ENG (respectivement, Λ de Wilks = 0,965 ; $p = 0,045$; $\eta^2 = 0,035$ et Λ de Wilks = 0,898 ; $p = 0,001$; $\eta^2 = 0,102$), indiquant que la qualité des informations fournies et l'engagement dans le cours à distance influencent l'acquisition des savoir-faire, notamment identifier des soutiens dans son environnement pour son projet et que le caractère engageant des activités joue un rôle déterminant dans l'appropriation de ce savoir-faire. Pour IESP, l'interaction avec l'engagement des étudiants (ENG) révèle également un effet significatif (Λ de Wilks = 0,956 ; $p = 0,026$), ce qui souligne que l'engagement influence aussi leur capacité à identifier des soutiens dans son environnement pour son projet « IESP ». En ce qui concerne les autres sous-dimensions, aucun effet significatif d'interaction avec les variables instrumentales n'a été observé.

Les préoccupations

En ce qui concerne les préoccupations, les résultats indiquent une amélioration significative de la dimension dans son ensemble ($t = 3,701$; $p < 0,01$). Les mêmes tests par sous-dimension indiquent que seule EOA montre une amélioration significative entre le pré-test et le post-test ($t = 3,70$; $p < 0,05$), ce qui suggère que la formation hybride a exercé un effet notable sur les choix d'orientation des étudiants. Parmi les résultats des tests multivariés, seuls les tests d'interaction entre EOA et INFO révèlent un effet significatif (Λ de Wilks = 0,958 ; $p = 0,029$, $\eta^2 = 0,042$). Cela indique une relation significative entre la qualité de l'information et l'évolution des choix des étudiants dans le cadre de l'orientation.

Le SEPE

Les résultats indiquent une amélioration significative pour SEPE ($t = 4,70$; $p < 0,01$). Cette différence souligne une progression notable des compétences des étudiants après la formation. En revanche, les tests multivariés n'ont révélé aucune interaction significative avec les variables liées à la perception de l'outil numérique.

6 Discussion et conclusion

Les résultats de l'enquête pré- et post-test montrent que les étudiants ont connu une progression dans des compétences à s'orienter, bien que certaines variables n'aient pas montré de changement statistiquement significatif. Pour les savoir-faire, seules deux

sous-dimensions liées aux savoir-faire en matière d'orientation ont progressé de manière significative. Cela indique un potentiel effet du dispositif hybride XXX, mais non démontrable, qui a permis aux étudiants d'acquérir des compétences en élaboration et mise en œuvre de plan d'action et en mobilisation et extension du réseau socioprofessionnel. Les tests multivariés indiquent que c'est plutôt spécifiquement deux autres dimension, « se découvrir et identifier ses compétences » (SDIC) et « identifier des soutiens dans son environnement pour son projet » (IESP) qui ont évolué en interaction avec des perceptions positives du dispositif, à savoir clarté des informations pour SDIC et IESP, et engagement en plus pour SDIC. Autrement dit, l'accent mis par les concepteurs sur la clarté et la pertinence des informations fournies dans le cours et l'interactivité et à la participation active des apprenants dans les activités proposées au sein du cours, ont permis ces progressions au sein de la dimension des savoir-faire.

En termes de préoccupation, c'est l'exploration des options alternatives, dans l'absolu et en interaction avec la perception de la clarté des informations, qui a évolué positivement. Ainsi, le dispositif hybride incite les étudiants à réfléchir à des alternatives professionnelles ou à des changements de parcours grâce à la manière dont les informations ont été mises à disposition.

Enfin, le sentiment d'efficacité personnelle a évolué de manière significative, mais pas en interaction avec la perception du dispositif : nous ne sommes donc pas en mesure d'indiquer qu'il s'agit d'un effet de la formation hybride ou bien simplement de l'évolution générale des étudiants entre le pré- et le post-test.

De manière intéressante, nous constatons que des variables de perception du dispositif de formation « se couplent » avec des variables d'apprentissage mais pas avec d'autres. Par exemple, pour les savoir-faire, les variables ayant progressé dans l'absolu ne sont pas celles qui ont progressé en interaction avec la perception instrumentale. Cela indique une hétérogénéité au sein de sous-dimensions appartenant à un même domaine d'apprentissage (ici les savoir-faire en termes d'orientation), en fonction de la perception de qualité d'un dispositif techno-pédagogique. Le système technique agit ainsi différemment sur différentes sous-dimensions d'une même variable. Ce qui peut expliquer que, si dans certaines études la recherche trouve de effets neutres des dispositifs à distance ou hybrides sur l'apprentissage, il se peut que des effets positifs sur des variables d'apprentissages, peut-être informels, soient existants mais non mesurés à travers les instruments utilisés. En effet, « des résultats non significatifs attestent avant tout que les différences [entre les résultats d'apprentissage dans un environnement EIAH et un environnement traditionnel] observées sont trop faibles pour être raisonnablement attribuées aux variables contrôlées par l'observateur, l'EIAH constituant bien souvent la variable la plus facile à contrôler » (Marquet, 2005, p. 386). Du point de vue dispositionnel, ce sont surtout des compétences pratiques d'étudiants en 2ème année qui ont évolué et non des préoccupations liées à la carrière professionnelle. Le caractère situé de notre étude nous incline à proposer que, si la même étude était faite avec des promotions plus « proches » du marché de l'emploi, ou alors avec une population déjà active, les résultats concernant les préoccupations seraient peut-être différents. Ainsi, la per-

ception d'un environnement EIAH ne peut dépendre seulement de ses propriétés pédagogique et de ses caractéristiques techniques. Si le chercheur reste « aveugle » quant aux réalités sociales des sujets étudiés, les résultats (qu'ils soient significatifs ou non, positifs ou non), ne sauraient s'expliquer uniquement par leur acceptabilité du système proposé.

Dans notre cas, l'analyse de la perception des qualités instrumentales du dispositif de formation hybride met en lumière plusieurs facteurs clés contribuant à la progression des étudiants. Par exemple, la qualité des informations transmises dans le cadre de la formation a influencé dans une grande mesure l'acquisition des savoir-faire nécessaires à leur orientation. L'engagement des étudiants à travers les activités proposées s'avère déterminant pour l'appropriation des savoir-faire, illustrant l'importance de l'interactivité dans le dispositif. Bien que certaines dimensions n'aient pas montré une évolution significative, l'ensemble des résultats conforte l'idée que l'intégration de dispositifs hybrides, alliant apprentissage en ligne et présentiel, peut jouer un rôle déterminant dans le développement des compétences à s'orienter des étudiants à l'université.

Références

1. Bandura, A. (1997). *Self-efficacy: The exercise of control*. W.H. Freeman.
2. Beltrame, F. (2019). *SavOrient : Savoir s'orienter au 21e siècle*. Creative Commons. Consulté à l'adresse http://deporientation.free.fr/SavORIENT/SavORIENT_001.php
3. Beltrame, F. (2022). Proposition méthodologique pour acquérir et développer les compétences à s'orienter. *Éducation Permanente, 232*, 135-144. https://doi.org/10.3917/edpe.232.0135
4. Caron, P.-A., & Heutte, J. (2017). Comprendre l'usage que les professeurs des écoles font des TNI et du numérique. In *EIAH 2017* (p. 476). Strasbourg, France: N. Guin, B. De Lièvre, M. Trestini, & B. Coulibaly (Eds.).
5. Flavell, J. H. (1987). Speculations about the nature and development of metacognition. In F. E. Weinert & R. Kluwe (Eds.), Metacognition, motivation, and understanding (pp. 1-29). Lawrence Erlbaum Associates.
6. Havighurst, R. J. (1953). *Human development and education*. Longmans, Green and Co.
7. Heath, D. H. (1960). *Growing up in college: Liberal education and maturity*. Jossey-Bass.
8. Heutte, J. (2011). La part du collectif dans la motivation et son impact sur le bien-être comme médiateur de la réussite des étudiants : Complémentarités et contributions entre l'autodétermination, l'auto-efficacité et l'autotélisme (Thèse de doctorat). Université Paris Ouest-Nanterre-La Défense, Nanterre.
9. Le Boterf, G. (2000). Construire les compétences individuelles et collectives : Agir et réussir avec compétence. Éditions d'Organisation.
10. Marquet, P. (2005, Mai). Intérêt du concept de conflit instrumental pour la compréhension des usages des EIAH. *Dans EIAH 2005 (Environnements Informatiques pour l'Apprentissage Humain)* (pp. 383-388). INRP.
11. Michaud, P. (2003). Compétences à s'orienter et développement professionnel. *Canadian Journal of Career Development, 2*(1), 21-30.

12. Michaud, G. (2003). Étude du transfert des compétences dans le cadre de démarche de counseling d'orientation (Thèse de doctorat, Université de Sherbrooke, Sherbrooke, Canada). http://savoirs.usherbrooke.ca/handle/11143/911
13. Molinari, G., Poellhuber, B., Heutte, J., Lavoué, E., Sutter Widmer, D., & Caron, P.-A. (2016). L'engagement et la persistance dans les dispositifs de formation en ligne : regards croisés. *Distances et médiations des savoirs*, *(13)*. https://doi.org/10.4000/dms.1332
14. Pôle emploi. (2013). *Rapport annuel sur l'état du marché du travail en France.* Paris: Pôle emploi.
15. Peraya, D., Charlier, B., & Deschryver, N. (2014). Une première approche de l'hybridation. *Education et formation*, (e-301), 15–34. http://archive-ouverte.unige.ch/unige:37049
16. Romainville, M. (2001). L'échec dans l'université de masse. *Revue française de pédagogie*, 136, 194-195
17. Schwarzer, R., & Jerusalem, M. (1995). Generalized self-efficacy scale. In J. Weinman, S. Wright, & M. Johnston (Eds.), *Measures in health psychology: A user's portfolio. Causal and control beliefs* (pp. 35-37). NFER-NELSON.
18. Super, D. E. (1955). *The theory of vocational development*. Harper & Brothers.
19. Poellhuber, B., & Roy, N. (2018). Satisfaction des adultes et perceptions de la qualité en formation donnée par les MOOC. Dans A. Jézégou, P.-A. Caron, & J. Heutte (Eds.), *Actes du Colloque e Formation des jeunes et des adultes 2018* (pp. xx-xx). Lille, France, mars 2018.

Scraping d'Environnements Numériques de Travail et Analyse de Curriculums : un Cas d'Etude

Matthieu Cisel[1], et Camille Bidaud[2]

[1] CY Cergy Paris Université, Laboratoire AGORA, Cergy, France
matthieu.cisel@cyu.fr
[2] ENSA Paris-Val de Seine, Laboratoire EVCAU, Paris, France
camille.bidaud@paris-valdeseine.archi.fr

Résumé. Les recherches sur les curriculums sont souvent limitées par le caractère manuel de la collecte de données. Mobiliser des méthodes issues de l'informatique pour automatiser la collecte et le traitement de données sur des contenus d'enseignement offre un certain nombre d'opportunités en termes de passage à l'échelle des analyses, considération que nous nous proposons d'illustrer avec le cas des Ecoles Nationales Supérieures d'Architecture. Via des techniques fondées notamment sur le scraping de données issues d'Environnements Numériques de Travail, nous nous penchons dans cette réflexion méthodologique sur l'emploi du terme « numérique », dans les intitulés des cours de onze établissements. Nous suivons les évolutions de l'emploi des termes sur plus d'une décennie, et contrastons leur fréquence selon les établissements.

Mots-clés : Environnements Numériques de Travail, textométrie, curriculums, analyse de contenu, scraping.

Abstract. Curriculum research is often limited by the manual nature of data collection. Automating through computer science the collection and processing of data on educational content offers numerous opportunities, a consideration that we propose to take into account. illustrate with the case of the French Higher Schools of Architecture. In this article focused on methodology, we scraped data from Virtual Learning Environments to describe the use of the term "digital", in the course titles of eleven establishments. We follow developments in the use of terms over more than a decade, and contrast the situations across institutions.

Keywords: Virtual Learning Environment, textometry, curricula, content analysis, scraping.

1 Introduction

En sciences de l'éducation, l'analyse des curriculums et de leurs évolutions constitue un champ de recherche investi depuis plusieurs décennies, tant dans la littérature francophone que dans la littérature anglo-saxonne [4,5,6]. Dans le domaine de

l'enseignement scolaire, les chercheurs se sont focalisés sur des objets comme les manuels, ou les programmes tels que publiés au journal officiel [10,11]. Dans l'enseignement supérieur, il existe un nombre considérable de travaux [12], mais les analyses se heurtent notamment à la grande diversité des formations qui caractérisent l'offre d'une université ou d'une grande école. Là où au niveau du collège ou de l'école primaire, le programme des sciences de la vie et de la terre est identique en tout point du territoire national, celui des formations de biologie cellulaire, par exemple, pose davantage de problèmes pour des analyses comparatives à grande échelle entre établissements.

L'étude des différences de curriculums qui existent entre institutions se heurte à l'hétérogénéité des documents sur lesquels sont présentés les programmes pédagogiques : plaquettes, sites internet aux structures variées, programmes au format PDF. Constituer manuellement une base de données sur la base de tels supports représente un travail laborieux, d'où l'intérêt que présente des stratégies d'automatisation d'une telle tâche. Pour ce faire, nous défendons dans cet article l'idée selon laquelle les Environnements Numériques de Travail (ENT) représentent pour l'étude des curriculums une alternative pertinente aux supports de communication destinés à faire la promotion des formations.

D'après la CNIL, un ENT représente « tout ensemble intégré de services numériques choisis et mis à disposition de tous les acteurs de la communauté éducative d'un ou plusieurs établissements de l'enseignement scolaire ou de l'enseignement supérieur […] l'ENT constitue un point d'entrée unifié permettant à l'utilisateur d'accéder, selon son profil et son niveau d'habilitation, aux services et contenus numériques dont il dispose » [1].

Dans le champ des EIAH, les ENT ont fait l'objet d'une littérature abondante, centrée essentiellement sur le secondaire [2]. A notre connaissance, il n'existe pas de travaux portant sur l'utilisation de ces outils pour étudier des curriculums, à plus forte raison à l'échelle de plusieurs établissements. Pourtant, ils constituent des objets de choix pour ce type d'analyse, et ce pour plusieurs raisons. En premier lieu, toutes les formations de l'établissement sont accessibles via la même technologie ; l'ENT permet donc de constituer une base sur l'ensemble des cours sur plusieurs années, ce qui ouvre la voie à des analyses diachroniques. En second lieu, les établissements d'enseignement supérieur mobilisent en France un nombre relativement réduit de technologies différentes (Moodle, le plus fréquent, Sakai, Blackboard, etc.). Cela signifie que si l'on développe pour une université une méthodologie pour constituer une base sur les enseignements – dans le cas fréquent où l'accès à une base de données brutes avec tous les cours serait impossible, l'on peut facilement l'extrapoler à de nombreuses autres universités mobilisant la même technologie. Se pose toujours la question des droits d'accès aux données, mais cet obstacle administratif peut être surmonté, à condition d'avoir dans les établissements d'intérêt des contacts de collègues disposant eux de ce droit d'accès, et qu'ils soient disposés à collaborer avec les responsables de l'étude.

Parmi les nombreuses opportunités que présente pour la recherche une base de données d'intitulés de cours, et fondée sur l'exploitation d'ENT, l'on peut nommer l'étude de l'évolution des termes mobilisés pour décrire les enseignements. Celle-ci permet de capturer des tendances, et de se pencher sur les liens entre les injonctions politiques, les

[1] https://www.cnil.fr/fr/definition/ent-espace-numerique-de-travail

débats dans la sphère publique et leurs traductions éventuelles dans les curriculums. Par ailleurs, l'on peut s'intéresser aux différences entre curriculums selon une logique d'analyse comparative.

Le présent article décrit une étude exploratoire visant à répondre à une question de nature méthodologique illustrée par un cas d'étude, les Écoles Nationales Supérieures d'Architecture (ENSA) : Dans quelle mesure l'analyse de curriculums constitués sur la base de l'extraction de données depuis des ENT permet-elle de visualiser des évolutions de la popularité de certains sujets dans les intitulés de cours d'une part, et d'appréhender des différences entre établissements d'autre part ?

Nous nous proposons de répondre à cette question sur la base de l'analyse de données issues de Taïga, en nous focalisant sur l'enseignement de matières liées au numérique. Taïga est un ENT développé exclusivement pour les ENSA, et utilisé dans l'intégralité de ces institutions. Ce faisant, nous illustrons comment des méthodes issues de l'informatique – et en particulier le scraping et l'analyse textométrique – peuvent être utilisées dans les recherches sur les curriculums. Celles-ci mobilisent traditionnellement des analyses de contenu largement manuelles, alors que l'automatisation au cœur de notre étude nous a permis de couvrir aisément l'évolution de plus d'une décennie d'intitulés de cours pour onze ENSA.

2 Méthode

2.1 Un Environnement Numérique de Travail riche en informations

L'ENT Taïga a déjà été mobilisé pour caractériser l'évolution d'enseignements en architecture, mais uniquement sur la base de l'utilisation de son moteur de recherche [3]. Celui-ci permet d'effectuer des requêtes par mots-clés, et donc d'obtenir des données chiffrées sur le nombre d'occurrences d'un terme en particulier dans les fiches pédagogiques associées aux cours. Utiliser directement un tel moteur présente plusieurs limites : en premier lieu, elle impose de réaliser des analyses année par année, et donc de compiler manuellement les résultats si l'on souhaite appréhender des évolutions au fil du temps. En second lieu, le moteur de recherche le terme d'intérêt que dans des fiches descriptives pédagogiques remplies de manière inégale entre écoles, et qui comprennent le titre de l'unité d'enseignement (U.E.), mais aussi parfois une partie dédiée à la bibliographie.

Ainsi, il suffit qu'une référence bibliographique contienne le mot « numérique » dans le syllabus détaillé d'une U.E. portant sur un sujet qui n'est pas nécessairement connecté pour que le décompte la prenne en compte au même titre que la création de nouveaux enseignements mobilisant effectivement le numérique. Dans le meilleur des cas, un nettoyage laborieux est nécessaire. Par conséquent, ne pouvant accéder au jeu de données brutes sur lequel repose le moteur de recherche, nous avons reconstitué une base de données en capitalisant sur notre accès aux ENSA.

Certaines personnes disposant d'un accès administratif à Taïga peuvent en effet, pour leur institution, extraire l'intégralité des enseignements, et ce jusqu'à l'année scolaire 2013-2014. Nous avons contacté des collègues des différentes ENSA pour obtenir

ces informations via des documents au format html ou csv. Un html est typiquement produit pour un niveau (licence ou master), pour une année donnée. Ce faisant, nous avons compilé les données de 11 ENSA, nommément Belleville, Bordeaux, Bretagne, La Villette, Lyon, Malaquais, Marne, Normandie, Nantes, Val de Seine, Toulouse. Les documents html reprennent l'ensemble des titres des U.E. et des différents enseignements. Les documents csv peuvent venir compléter la base, après jointure, lorsque les enseignements sont peu détaillés dans l'html.

2.2 Choix des termes constituant la focale de l'étude

Nous avons ensuite compté les occurrences du terme « numérique » dans les enseignements, selon une méthodologie que nous précisons par la suite, après avoir exploré d'autres termes comme « digital », et « ordinateur ». Ce terme a été choisi suite à l'exploration des données. Une exploration préliminaire de termes synonymes ou appartenant au même champ sémantique, tels que *digital* ou *ordinateur*, n'a en effet pas permis d'observer de tendance significative : leurs occurrences étaient trop rares pour fournir des résultats exploitables ou pour révéler une évolution comparable. En ce sens, le terme *numérique* apparaît comme le plus représentatif de mutations des enseignements dans le contexte francophone, notamment pour suivre les évolutions discursives et institutionnelles autour de l'enseignement de l'architecture.

S'agissant du choix du terme au centre de l'étude, il convient de noter que l'une des difficultés posées par la focalisation sur les intitulés de cours réside dans le fait qu'il est difficile d'identifier si des évolutions – ou des différences entre établissements – relèvent de modifications substantielles des maquettes pédagogiques ou s'il s'agit avant tout de nuances terminologiques. Il peut y avoir simplement substitution des termes dans les curriculums, résultat constaté dans un autre article pour le terme *transition*, qui a remplacé peu à peu le terme *durable*, présent notamment dans *développement durable* [1]. Or une telle évolution terminologique reflète sans doute une question de tendance, de « buzzword », qu'il est difficile de distinguer, par ces méthodologies, d'un changement significatif dans les contenus présentés aux étudiants.

En théorie, une évolution du nombre des enseignements avec l'intitulé *numérique* peut aussi bien refléter une fluctuation, à l'échelle nationale, de la priorité donnée à la maîtrise d'outils digitaux dans les cours des ENSA, que de changements dans la manière de décrire les cours. Néanmoins, en recherchant *digital*, *informatique*, *ordinateur*, *modèle*, aucun terme ne semble être très utilisé, ou compenser d'évolution de *numérique*, contrairement au phénomène de vase communicant entre *durable* et *transition*. Il existe des termes spécifiques à l'architecture comme « AutoCAD », logiciel dédié à la conception en 3D, mais seulement de manière anecdotique. L'expérience acquise par l'un des co-auteurs suggère que dans les cours en architecture, les étudiants peuvent être amenés à utiliser des outils « numériques » dans des enseignements centrés, dans leur intitulé, sur la géométrie, sans qu'aucun emploi d'un synonyme de *numérique* ne puisse déceler une telle pratique. C'est là la limite des analyses par intitulé de cours.

En se contentant de l'intitulé de cours, l'on ne peut pas voir des évolutions en profondeur d'un curriculum. Mais comprendre l'évolution en profondeur du contenu d'un cours n'est pas possible à l'échelle d'une décennie, sur l'intégralité des cours d'une

dizaine d'écoles. Un décompte de mots-clés permet de construire une première série d'indicateurs. C'est une première étape pour appréhender des évolutions en cours, qu'il faut bien évidemment approfondir par la suite pour éviter les erreurs d'interprétation.

2.3 Collecte et analyse des données

R 4.0 a été mobilisé tant pour extraire des fichiers html et csv, par scraping (librairie *rvest*) le contenu d'intérêt que pour l'analyse textométrique qui a suivi. R est un langage de programmation généraliste particulièrement pertinent pour les statistiques, mais aussi efficace pour les tâches simples de scraping, au même titre que Python.

Le scraping (ou moissonnage de données) désigne l'ensemble des techniques permettant d'extraire automatiquement des informations accessibles sur le web ou dans des bases de données [7]. En contournant l'accès manuel, le scraping facilite la collecte massive de contenus. Il faut interagir avec la structure HTML d'une page pour en extraire textes, liens, métadonnées, etc. Lors d'une opération de scraping, le choix des balises HTML est crucial : ce sont elles qui structurent l'information. Sélectionner les bonnes balises permet de cibler précisément les contenus pertinents et d'éviter de récupérer du bruit (menus, bas de page). Toutes années confondues, nous avons collecté par cette voie des données sur 10807 enseignements en licence et 20915 en master.

Une fois les données collectées, l'analyse textométrique offre des outils pour explorer systématiquement les contenus textuels. Issue de la linguistique quantitative et de l'analyse de corpus, la textométrie vise à objectiver l'étude des textes en quantifiant leur structure et leur lexique. L'un des indicateurs de base est la fréquence des termes (term frequency, TF), qui mesure combien de fois un mot apparaît dans un corpus donné, en l'occurrence une base de données d'intitulés de cours, et qui permet de produire des vecteurs de mots. Ce comptage simple, fait avec R au travers d'une approche de type « expression régulière », permet de repérer les thématiques dominantes, ou l'évolution d'une thématique d'intérêt. De là, l'on peut identifier les mots les plus saillants, ou encore de construire des représentations visuelles comme des nuages de mots ou des cartes de cooccurrence. Nous avons mis en place une approche empêchant des redondances : le décompte du terme « numérique » est fait sur la seule base de l'intitulé des cours, et ne prend pas en compte une éventuelle apparition dans d'autres éléments descriptifs (tels que les ressources bibliographiques), contrairement à ce qui a pu être fait dans des études antérieures [3].

Le calcul de la TF est souvent utilisé seul ou combiné avec d'autres métriques, comme le TF-IDF (term frequency–inverse document frequency), qui pondère la fréquence d'un mot par sa rareté dans l'ensemble du corpus, permettant ainsi de distinguer les termes véritablement spécifiques d'un texte [8]. C'est par exemple ce type de métriques qui a utilisé par Fortino *et al.* [9] pour mesurer l'adéquation entre les termes mobilisé dans un curriculum (intitulés de cours) et les titres de postes auxquels postulent les étudiants de différentes formations d'universités américaines. Néanmoins, dans la mesure où nous souhaitons comparer l'évolution d'un terme donné au fil des ans, la mobilisation de la métrique TF a semblé plus pertinente ici.

3 Résultats

Dans cette section, nous présentons d'une part une analyse diachronique de l'emploi du terme « numérique », tous établissements confondus, et d'autre part une comparaison entre établissements entre licences et masters. En Figure 1, nous constatons un emploi croissant du terme tant en licence qu'en master, avec un pic autour de 2020 correspondant à plus de soixante-dix occurrences, suivie d'une baisse progressive et limité de son emploi. Il y a un décalage d'un an dans les évolutions entre masters et licences.

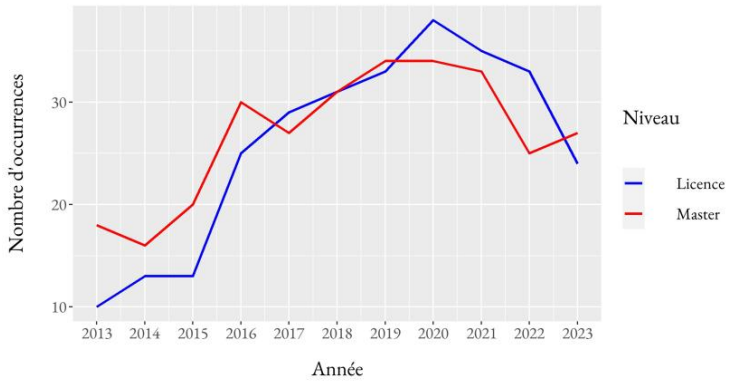


Fig. 1. Analyse diachronique de l'utilisation du terme « numérique » dans les intitulés d'enseignements de 11 ENSA de 2013 à 2023. On notera que 2013 correspond à l'année 2013-2014

L'analyse permet également de visualiser des différences entre établissements (Figure 2). On constate par exemple que les ENSA Bretagne et La Villette n'emploient pas le terme « numérique » dans les cours de licence. Pour certains établissements comme l'ENSA Bordeaux, le nombre d'occurrences est supérieur pour le niveau licence, pour d'autres, comme Val de Seine, il est supérieur au niveau master, reflétant la place que les instances décisionnaires veulent voire le numérique occuper dans les enseignements dans la séquence de cours.

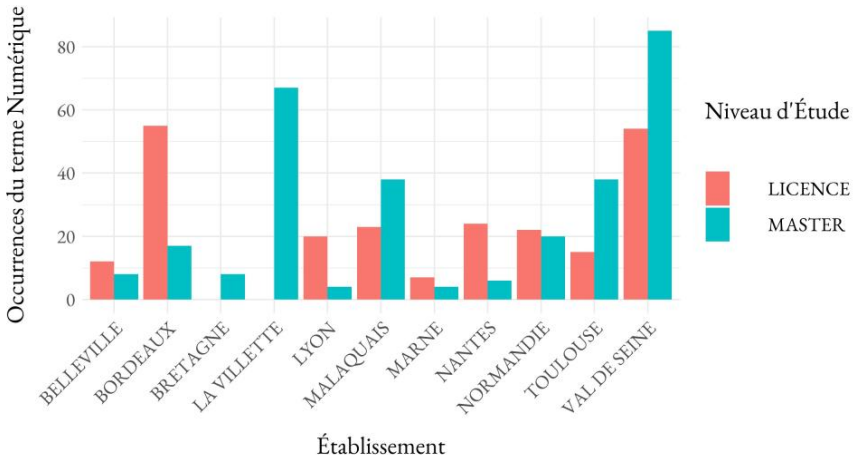


Fig. 2. Utilisation, différenciée selon le niveau et l'établissement, du terme « numérique » dans les intitulés d'enseignements de onze ENSA de 2013 à 2023.

4 Discussion

L'expérience d'enseignement acquise par l'un des co-auteurs à La Villette et dans d'autres ENSA suggère que l'absence ou la faible représentation du terme *numérique* en licence ne découle pas d'une simple évolution terminologique. C'est davantage le reflet d'un choix délibéré des établissements de retarder le plus possible l'utilisation des outils numériques dans le cursus de l'école. Dans la mesure où il s'agit d'une preuve de concept d'une approche méthodologique, nous ne souhaitons pas ici analyser plus avant l'historique de l'utilisation d'un terme comme *numérique*, ni les déterminants des évolutions observées. Pour un travail centré sur l'interprétation des tendances observées, nous renvoyons à un article basé sur la présente base de données, et consacré à l'enseignement de la réhabilitation [1], et aux recherches consacrées aux déterminants des évolutions [4].

Le recours au scraping des Environnements Numériques de Travail (ENT) pour collecter les intitulés de cours constitue une approche méthodologique que nous pensons innovante pour la conception d'indicateurs relatifs à l'évolution des curricula. L'extraction systématique de ces intitulés permet de construire des indicateurs empiriques, fondés sur la fréquence d'occurrence de certains termes (par exemple numérique), témoignant de l'intégration progressive de thématiques émergentes dans les formations, ou d'évolutions dans les termes utilisés. Toutefois, plusieurs limites doivent être soulignées. D'abord, les intitulés captés ne garantissent pas une exhaustivité parfaite : certaines informations peuvent être absentes, obsolètes ou partiellement mises à jour dans

les ENT. Ensuite, comme nous l'avons souligné, l'intitulé seul ne renseigne qu'indirectement sur les contenus effectifs des enseignements, exposant l'indicateur à un risque de surinterprétation. Enfin, l'hétérogénéité des pratiques d'intitulé selon les établissements ou les enseignants introduit une variabilité qui doit être méthodologiquement contrôlée (par normalisation ou codage thématique). Malgré ces limites, le scraping des ENT offre une opportunité rare d'observer, avec une granularité temporelle et institutionnelle fine, la dynamique d'institutionnalisation de nouveaux savoirs, et de documenter les évolutions discursives de l'offre pédagogique.

Toute conception d'indicateur est confrontée à des limites consubstantielles, qui tiennent à la nature même de l'opération de mesure. Un indicateur ne capture qu'une traduction partielle et construite du phénomène étudié : il repose nécessairement sur des choix de définition, de périmètre et de catégorisation qui introduisent un écart entre la complexité du réel et sa représentation. Par ailleurs, tout indicateur simplifie les dynamiques sous-jacentes en les ramenant à des éléments quantifiables, parfois au prix d'une perte de sens contextuel ou qualitatif.

Dans cette perspective, l'indicateur, et les représentations graphiques que nous en avons faites, doivent être compris non comme un aboutissement, mais comme une première étape dans l'analyse : ils permettent de repérer des phénomènes émergents, des inflexions ou des tensions qu'il s'agit ensuite de creuser par des méthodes complémentaires (enquêtes qualitatives, analyses de contenu, observations de terrain).

5 Conclusion

Si notre approche fondée sur les intitulés de cours permet de capturer certaines tendances à l'œuvre, elle reste superficielle dans la mesure où elle ne permet pas nécessairement d'appréhender d'évolutions significatives quant aux contenus enseignés, puisqu'elle n'entre pas dans le détail du contenu des enseignements. Les fiches pédagogiques contiennent de nombreux autres éléments qui peuvent également être exploités, à condition d'une homogénéisation des pratiques de remplissage de ces fiches. La bibliographie recommandée aux étudiants constitue par exemple une rubrique présentant davantage d'informations sur le contenu des enseignements que le simple intitulé du cours. Une analyse à grande échelle d'une base focalisée sur les seules références fournies aux étudiants permettrait ainsi de réaliser des analyses plus approfondies.

6 Références

1. Bidaud, C., et Cisel, M. (2024). Une décennie d'essor des enseignements sur l'existant dans les ENSA vue à travers les intitulés de cours. *Les Cahiers de la recherche architecturale urbaine et paysagère*, *22*. https://doi.org/10.4000/13354
2. Bruillard, É. (2011). Le déploiement des ENT dans l'enseignement secondaire: entre acteurs multiples, dénis et illusions. *Revue française de pédagogie*, *177*, 101-130.
3. Cremnitzer, J.-B. (dir.), Valter Balducci (coord*.), Former à la réhabilitation, opus cit.,* p. 14
4. Krause, K.-L. D. (2022). Vectors of change in higher education curricula. *Journal of Curriculum Studies*, *54*(1), 38–52. https://doi.org/10.1080/00220272.2020.1764627

5. Lemaître, D. (2009). Le curriculum des grandes écoles en France: Un modèle d'analyse inspiré de Basil Bernstein. *Revue française de pédagogie. Recherches en éducation*, *166*, Article 166. https://doi.org/10.4000/rfp.1096
6. Martinand, J.-L. (2014). Point de vue V – Didactique des sciences et techniques, didactique du curriculum. *Éducation et didactique*, *8*(1)
7. Mitchell, R. (2024). *Web Scraping with Python: Collecting Data from the Modern Web*. *3rd Edition.* O'Reilly Media.
8. Ramos, J. (2003, Janvier). Using TF-IDF to Determine Word Relevance in Document Queries. Dans *Proceedings of the First Instructional Conference on Machine Learning*. (pp 24-37).
9. Fortino, A., Zhong, Q., Huang, W. C., & Lowrance, R. (2019, Mars). Application of text data mining to stem curriculum selection and development. Dans *2019 IEEE Integrated STEM Education Conference (ISEC)* (pp. 354-361). IEEE.
10. Denizot, N. (2016). Le manuel scolaire, un terrain de recherches en didactique?. *Le français aujourd'hui*, *194*(3), 35-46.
11. Bruillard, É. (dir.) (2005). *Manuels scolaires, regards croisés*. Caen : CRDP de Basse-Normandie.

Évaluation des réponses courtes à l'aide de la similarité de phrases

Michel C. Desmarais[1][0000−0002−2990−1078], Ovide Bertrand Kuichua Kamdem[1], and Arman Bakhtiari[1]

Polytechnique Montréal, Montréal, Canada
prenom.nom@polymtl.ca

Résumé Cet article propose une approche à l'évaluation automatique des réponses courtes (Automatic Short Answer Grading, ASAG) basée sur l'utilisation d'encodeur de phrases pour le calcul de similarité des réponses et une fonction de régression non linéaire pour la prédiction basée sur quelques réponses évaluées. Le contexte d'utilisation est celui un correcteur qui évalue manuellement une quinzaine de réponses, tandis que le reste est noté automatiquement. Ce contexte est relativement fréquent et correspond aux situations où un certain nombre de réponses d'élèves doivent être examinées avant que le correcteur soit suffisamment à l'aise pour attribuer une note finale. Les résultats montrent que l'approche proposée atteint une précision de notation comparable à celle des modèles de transformeurs affinés (*fine-tuned)*, qui sont beaucoup plus gourmands en calcul et plus complexes à utiliser pour l'entraînement, ce qui les rend moins attrayants pour de nombreux contextes d'utilisation. L'approche proposée a l'avantage d'être particulièrement facile et efficace à déployer, tout en offrant une bonne précision.

Mots clés : ASAG, Notation automatique des réponses courtes, évaluation des apprenants, similarité de textes, plongements de textes

Abstract. This paper proposes an approach to Automatic Short Answer Grading (ASAG) based on the use of sentence encoders for answer similarity computation and a nonlinear regression function for prediction based on a few evaluated answers. The context of use is that a grader manually grades about fifteen answers, while the rest are graded automatically. This context is relatively common and corresponds to situations where a certain number of student answers must be reviewed before the grader is comfortable enough to assign a final grade. The results show that the proposed approach achieves a grading accuracy comparable to that of fine-tuned transformer models, which are much more computationally intensive and complex to use for training, making them less attractive for many use cases. The proposed approach has the advantage of being particularly easy and efficient to deploy, while providing good accuracy.

Keywords : ASAG, Automatic Short Answer Grading, Learner evaluation, Text Similarity, text embeddings

1 Introduction

Comme toute application informatique, la conception d'un outil d'aide à la correction doit tenir compte du contexte d'utilisation. Ainsi, un contexte typique de correction d'un examen est qu'il comporte quelques dizaines de copies. Le correcteur corrige initialement une partie des copies afin d'affiner les critères de correction et "calibrer" la sévérité. Cette étape est particulièrement importante si l'enseignant doit fournir une ligne directrice à d'autres correcteurs qui l'assistent. La majorité des questions ne sont pas tirées d'anciens examens et la tâche implique quelques dizaines de copies à noter. On ne peut donc pas se reposer sur un corpus étendu de réponses antérieures.

Ce contexte constitue le cadre général pour l'approche de correction automatique des réponses courtes proposée dans ce texte, une tâche connue en anglais sous le nom de *Automatic Short Answer Grading, ASAG*. Contrairement à plusieurs approches populaires en ASAG, ce contexte ne se prête pas particulièrement bien à l'entraînement de modèles de langue, notamment à l'affinage fin (*fine-tuning*) qui nécessite un volume relativement important de données. Par exemple, avec les données que nous utilisons, Zhu et al. [23] affirme que "l'ensemble de données Mohler ne contient que 2273 paires de réponses, ce qui est trop peu pour les modèles d'apprentissage profond" et ils ont dû recourir une technique d'enrichissement des données pour amener ce jeu de données à 3300 réponses. De plus, l'affinage fin de LLM demeure relativement coûteux en temps de calcul. Ces exigences rendent l'approche d'affinage fin moins adaptée au contexte d'utilisation de notre étude.

Par contre, ce contexte se prête très bien à une approche de similarité de texte entre une réponse étudiante et la réponse de référence. Les grands modèles de langue (LLM) préentrainés sont rapides et relativement efficaces pour effectuer un tel calcul de similarité. Nous proposons une approche qui utilise un transformeur récent et un modèle de régression non linéaire simple pour calculer le score à chaque question. La régression est paramétrée à partir d'une quinzaine de réponses que le correcteur doit fournir au préalable. C'est donc une approche semi-automatisée que nous avançons, mais qui permet d'assister un enseignant dans la correction dans la mesure où les scores calculés sont fiables.

Les résultats obtenus avec un corpus de données très utilisé pour évaluer la performance de modèles pour ASAG, les données de Mohler [16], se comparent en termes de précision à ceux l'état de l'art obtenus avec des modèles de transformeurs affinés. L'approche proposée est particulièrement simple à développer et repose sur des calculs largement plus efficaces en termes temps que les modèles affinés.

2 Travaux portant sur la correction automatique

Les dernières années ont été particulièrement fertiles dans le domaine de l'ASAG grâce aux avancées des grands modèles de langage, les LLM qui reposent sur l'architecture de transformeur [21].

La très grande majorité des des travaux récents se reposent sur les modèles de langage basés sur l'architecture transformeurs. Ces approches adoptent souvent un modèle pré-entrainés pour la vectorisation de texte. En vectorisant le texte de la réponse attendue et de celle de l'étudiant, un calcul de la similarité sémantique fournit un indicateur de la validité de la réponse [7–9, 11]. Amur et al. [3] ont effectué une revue des nombreuses études qui ont adopté cette approche pour le calcul de la similarité de texte en général, dont les applications pour l'ASAG. L'approche proposée ici s'inscrit dans cette catégorie, avec de meilleurs résultats qui se comparent plutôt à ceux des approches qui adoptent l'affinement.

En effet, plusieurs études récentes utilisent une approche d'affinement (*fine-tuning*) des modèles transformeurs pour la classification [10, 12, 13, 23]. Gard et al. [12] utilisent des paires réponses, la réponse attendue et la réponse étudiante, pour l'affinement du transformeur et obtiennent un résultat qui dépasse les autres approches avec une erreur quadratique moyenne (EQM) de 0.732 pour les données de Mohler [16] que nous utilisons dans la présente étude. Zhu et al. [23] effectuent un affinement avec une petite partie des réponses étudiantes pour obtenir le meilleur score de corrélation Pearson avec les mêmes données, 0.897 [1].

Parmi les plus récents travaux, on retrouve ceux qui utilisent l'IA générative pour demander un score directement à l'IA. Les résultats de Chang et al. [6] et de Grévisse [14] démontrent une nette amélioration entre les versions précédentes des IA génératives comme ChatGPT, ce qui suggère que cette approche offre une avenue prometteuse, mais les auteurs de ces deux études s'entendent sur la nécessité de maintenir une supervision de l'évaluateur. Tobbler et al. [19] présentent un outil basé sur l'IA générative conçu pour noter les réponses des étudiants en les comparant à des réponses de référence. D'autre part, nous n'avons pas identifié d'études permettant de comparer les performances de cette approche et celles alternatives. En particulier, il n'existe pas à l'heure actuelle de score de performance de l'IA générative pour les données de Mohler qui permettrait une comparaison. Néanmoins, il semble acquis que l'IA générative puisse fournir une rétroaction utile [2] et qu'elle soit aussi utilisable pour enrichir des approches alternatives [5, 17, 22].

Mentionnons finalement l'étude de Agarwal et al. [1] qui utilise une approche d'apprentissage profond avec graphes et obtient de très bons résultats d'erreur quadratique de 0.76 avec les données de Mohler. L'approche nécessite cependant la création d'une structure de graphe basée sur l'analyse sémantique des phrases (AMR, [4]) et une phase d'entraînement avec GPU qui est vraisemblablement peu adaptée au contexte d'utilisation de notre étude.

---

1. À l'instar de plusieurs autres études, nous omettons ici celle de [20] dont les résultats pour les données Mohler sont de 0.949 pour la corrélation et de 0.04 pour l'erreur quadratique. Bien qu'impressionnants, ces résultats ne semblent pas comparables. Car, comme la corrélation des scores entre les deux correcteurs de ce jeu de données est de 0.597 et qu'un calcul de l'EQM interjuge donnerait 1.7, un facteur de 42 (1.7/0.04), les performances de 0.949 et 0.04 sont peu vraisemblables et nous présumerons qu'une erreur s'est glissée dans l'article.

Ce survol des travaux portant sur la correction automatique démontre que des percées importantes ont été réalisées dans la dernière décennie, surtout grâce aux grands modèles de langage. Cependant, les meilleures performances sont obtenues à partir de modèles qui nécessitent une phase de préentrainement coûteuse et qui s'avèrent peu adaptés au contexte d'utilisation de notre étude.

3 Évaluation par similarité sémantique et lissage gaussien d'un échantillon de réponses

Comme plusieurs autres avant elle, l'approche proposée repose principalement sur le calcul de la similarité sémantique avec des plongements obtenus par des transformeurs, mais elle se distingue par la présupposition que le correcteur fournira au préalable quelques réponses scorées pour chacune des questions pour guider le processus de correction. Cette contrainte semble raisonnable du fait que cette étape permet de mieux jauger la difficulté de la question et d'établir ainsi un calibrage de la rigueur de la correction. Elle permet aussi de vérifier au préalable si la réponse anticipée est bien la seule bonne réponse avant de procéder à une correction.

L'approche proposée consiste donc à simplement calculer la similarité cosinus entre les plongements des réponses attendues-étudiantes et à utiliser les scores observés pour définir une fonction non linéaire du calcul de la note sur une échelle de [0–5]. Une fonction non linéaire s'avère utile ici comme on peut constater plus loin à la question 1.4 de la figure 2 où des scores de 5/5 sont observés à partir d'une similarité supérieure à 0.6, alors que sous ce seuil la variabilité est plus forte. Une fonction linéaire influencerait négativement les scores des réponses dont la similarité est supérieure à 0.6 alors qu'une fonction non linéaire sera moins sensible aux erreurs créées par ce type de distribution.

Pour définir cette fonction, un lissage gaussien (donc non linéaire) est calculé avec les données disponibles pour définir un ensemble de points de prédiction. Une interpolation linéaire est ensuite faite pour déterminer la valeur finale d'une mesure de similarité entre deux points. À noter qu'il serait aussi possible de définir une fonction de régression gaussienne, mais l'approximation par interpolation s'avère efficace et efficiente, et congruente avec le principe de simplicité promu avec l'approche proposée.

D'autre part, il importe de remarquer qu'une courbe de lissage est calculée par question. En effet, la distribution des scores est très différente d'une question à l'autre (voir la section 4 qui suit et la figure 2). C'est du moins le cas pour les données de Mohler et il n'y a aucune raison de croire que ce soit une exception. Cette observation motive la création d'un modèle par question plutôt qu'un modèle général pour l'ensemble de l'examen.

4 Données

Le jeu de données Mohler [16] est une référence pour l'évaluation des systèmes de correction automatique des réponses courtes (ASAG) [3,11]. Il se compose de

80 questions et de 2 273 réponses d'étudiants, chacune notée sur une échelle de 0 à 5 par deux enseignants. Les données proviennent d'un cours d'introduction à l'informatique d'une université au Texas. La longueur moyenne des réponses de référence est de 15 à 20 mots. Quelques exemples sont fournis au tableau 1.

À noter que nous considérons que chaque question de ce corpus possède une et une seule bonne réponse. Il peut y avoir une certaine flexibilité dans formulation de la réponse, comme c'est le cas pour la première question du tableau 1 portant sur le rôle d'un prototype, mais nous considérons néanmoins ici qu'il n'y a qu'une seule bonne réponse. Une approche plus sophistiquée s'imposerait s'il devait y avoir plusieurs bonnes réponses.

Tableau 1. Exemples de questions-réponses du corpus de Mohler [16]

Question	**Réponse**	
	de référence	**de l'étudiant**
What is the role of a prototype program in problem solving ?	To simulate the behaviour of portions of the desired software product.	High risk problems are address in the prototype program to make sure that the program is feasible. A prototype may also be used to show a company that the software can be possibly programmed.
How many constructors can be created for a class ?	Unlimited number.	as many as the user decides to use.
How do you delete a node from a binary search tree ?	Find the node, then replace it with the leftmost node from its right subtree (or the rightmost node from its left subtree).	You must first traverse the tree to find the appropriate value. Then you must make sure that the node is a leaf node. If it is, then you can delete the pointer to that specific node.
What are the main operations associated with a stack ?	push and pop	pop, push

La figure 1 illustre la distribution du score des 2 273 réponses selon leur similarité avec leur réponse attendue respective. On constate qu'il y a un grand nombre de réponses qui sont de 5/5 et relativement peu de réponses sous le score de 4/5. Cependant, la distribution par question est fort variable comme on le voit à la figure suivante.

La figure 2 illustre les distributions du score par similarité pour quelques questions individuelles et d'environ 30 étudiants. Comme on le constate, ces distributions varient considérablement d'une question à l'autre. Par exemple, les questions 1.2 et 1.3 comportent un faible taux de corrélation, contrairement à

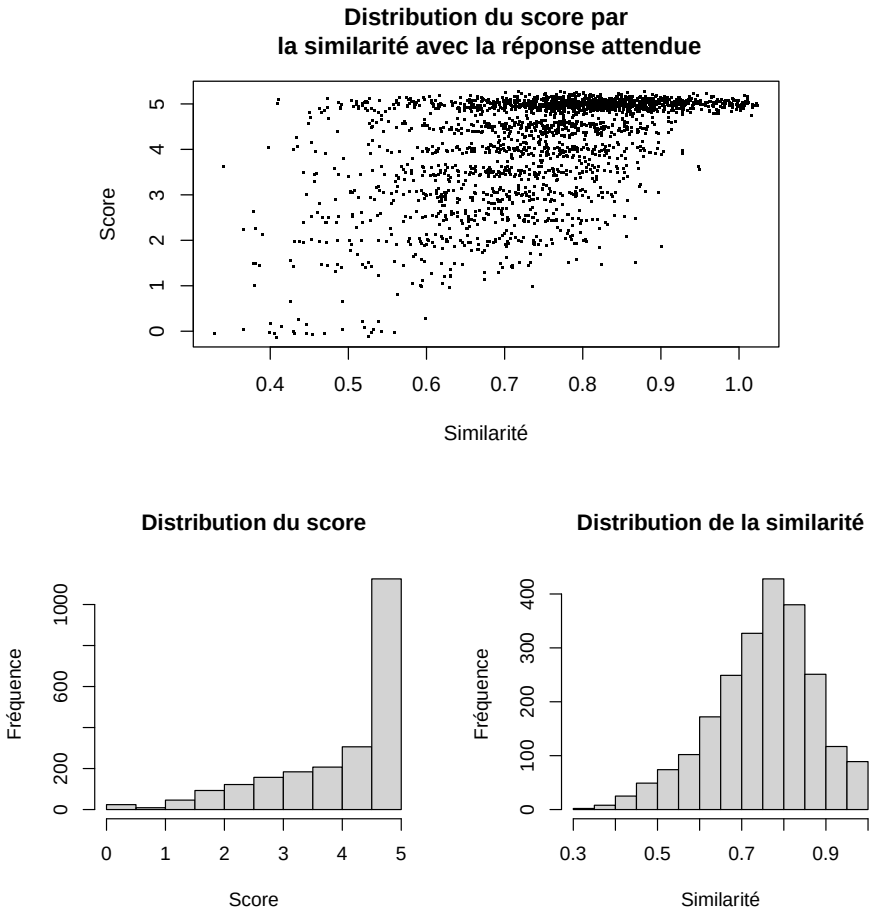


Figure 1. Figure du haut : distribution du score par la similarité entre la réponse de l'étudiant et celle attendue. Un bruit gaussien a été appliqué afin de discerner les points qui se superposent. Figure du bas : histogrammes des valeurs du score et de la similarité.

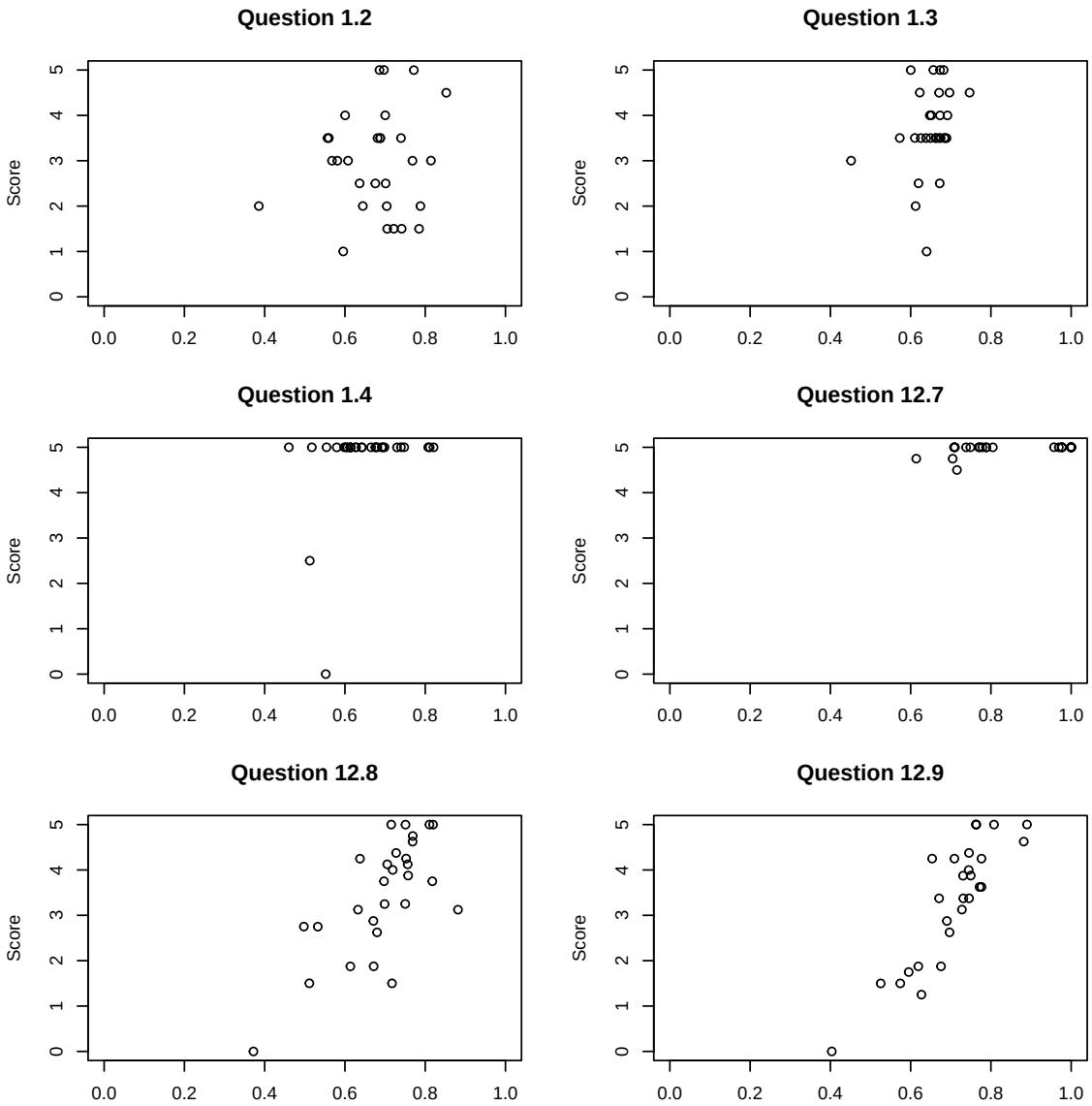


Figure 2. Distribution du score par la similarité (en x) de quelques questions.

12.9. Les questions 1.4 et 12.7 ont, quant à elles, un très haut taux de réussite qui rend le calcul de la corrélation très variable sur la base d'un échantillonnage. Compte tenu des particularités de chaque distribution, il s'avère plus efficace de modéliser une distribution propre à une question que pour l'ensemble.

5 Méthodologie

La tâche consiste à calculer un score à partir d'une mesure de similarité. Chacune des 2 273 réponses étudiantes est associée à une réponse attendue. Nous avons utilisé un modèle de langue pour obtenir un encodage pour chaque réponse sous la forme d'un vecteur dont la longueur varie typiquement de quelques centaines à quelques milliers de chiffres pour les plus grands modèles. Nous effectuons par la suite un calcul du cosinus entre les vecteurs qui sert de mesure de similarité. Aucun prétraitement des données n'est réalisé.

Le tableau 2 rapporte différentes statistiques des modèles pour l'encodage de phrases que nous avons étudiés. Notre choix s'est porté sur `mxbai-embed-large` dont la corrélation entre les similarités et les scores est la plus élevée. Ce modèle est performant. Il encode une centaine de questions en légèrement plus d'une seconde avec un processeur M4 par Apple.

L'objectif est d'obtenir une fonction de régression fournissant un score prédit pour une valeur de similarité entre la réponse étudiante et la réponse attendue. Cette fonction est repose sur un échantillon de 15 réponses pour lesquelles un score est obtenu par le correcteur. Plusieurs variantes ont été testées avec des résultats relativement semblables. La version utilisée ici consiste à calculer une somme pondérée des 15 scores fournis en fonction d'une distance Gaussienne de similarité avec la réponse pour laquelle on désire prédire un score. Une fonction est définie pour chaque question.

Le choix des 15 réponses est basé sur un échantillonnage uniforme des scores : une fois les réponses ordonnancées selon la similarité, la sélection est faite sur les rangs impairs.

Le code et les données utilisées l'expérimentation sont accessibles du lien `https://osf.io/rsg2w/?view_only=d5e180fe1463474f894e4065e3356b99`.

Tableau 2. Corrélations entre les réponses étudiantes et de référence pour trois modèles. Les autres informations rapportées sont le nombre de paramètres (en millions), la quantification, la taille des plongements (*embeddings*), le temps de traitement moyen pour 100 questions et l'année d'introduction.

Modèle	Corrélation	# Paramètres (m)	Quan.	Vec.	Temps	Année
all-minilm-l6-v2	0.482	22.6	–	384	0.5s	2021
bge-m3	0.517	576.6	16	1024	1.2s	2024
mxbai-embed-large	0.523	334	16	1024	1.2s	2024

6 Résultats

Le tableau 3 rapporte la performance du modèle proposé, simsemlis, en termes de corrélation et d'erreur quadratique moyenne (EQM). Ces résultats sont mis en comparaison avec ceux des approches de l'état de l'art [1, 12, 17, 23] et de deux autres modèles basés ceux-ci sur des approches plus classiques et dites *bag-of-words* [15, 18]. Les meilleures performances sont indiquées en caractère gras.

Tableau 3. Performance des modèles. Voir [1] pour une liste plus exhaustive de comparaisons ([1], tableau 1).

Approche	**Type**	**Corrélation**	**EQM**
simsemlis (notre approche)	Sim. sém.	0.623	0.755
État de l'art			
Garg et al. (2022) [12]	Rég. BERT + Sém. sim.	0.777	**0.732**
Agarwal et al. (2022) [1]	Graphe + transformeur	nd	0.762
Ouahrani et al. (2024) [17]	IAG+Sim. sém.	0.735	0.779
Zhu et al. (2022) [23]	BERT + LSTM.	**0.897**	0.827
Approches classiques			
Sultan et al. (2016) [18]	BOW	0.592	0.887
Mohler et al. (2011) [15]	BOW	nd	0.999

L'approche proposée se situe donc au deuxième rang des meilleures approches en termes d'EQM, mais au dernier ou avant dernier rang quant à la corrélation parmi les approches de l'état de l'art.

7 Discussion

On constate que les meilleurs performances diffèrent selon qu'on retient la corrélation ou l'EQM. Comme on peut le voir à la figure 2 des questions comme 1.4 et 12.7 peuvent avoir une corrélation faible, car presque tous les scores sont de 5/5, mais l'erreur quadratique peut être faible si la fonction du score a une courbe qui reflète bien la distribution. C'est particulièrement le cas pour la question 12.7 où la fonction de lissage gaussien. C'est aussi le cas pour 1.4, mais on remarque que deux réponses sont faibles (2.5 et 0) et que dans cette même zone de similarité on retrouve aussi des scores de 5.

La question se pose donc si c'est l'EQM ou la corrélation qui devrait être retenue comme mesure de performance étant donné l'écart relativement important du rang des meilleures approches. Étant donné que les distributions de scores par question s'éloignent fortement de la distribution gaussienne (ou normale)

et qu'il y a parfois un grand nombre de scores égaux (surtout pour la valeur de 5) et que ce n'est pas le cas des scores estimés, le choix de l'EQM s'impose. Nous avançons en fait que la corrélation seule n'est pas une mesure adéquate puisqu'elle est invariable à la translation et à la mise l'échelle. Par exemple, en supposant des valeurs observées de $[0, 1, 2]$, des valeurs prédites de $[3, 4, 5]$ donneraient une corrélation de 1 qui omet l'erreur de biais. Des valeurs prédites de $[4.0, 4.1, 4.2]$ (une mise à l'échelle) donneraient quant à elles une corrélation de 1. On voit bien ici comment la corrélation peut s'avérer trompeuse et c'est pourquoi nous adopterons uniquement l'EQM. De plus, l'EQM reflète davantage ce que l'utilisateur désire : un faible écart entre les valeurs prédites et observées.

En ne retenant que l'EQM, la méthode simsemlis se situe au deuxième rang de l'état de l'art. Compte tenu qu'elle est beaucoup moins exigeante en temps de calcul et en simplicité que les approches d'affinage de transformer [1, 12, 23], elle semble donc un choix valable.

Mentionnons cependant que l'approche de paraphrasage de la réponse attendue [17] se compare probablement à simsemlis en termes de simplicité puisqu'elle ne nécessite pas d'affinage. Le principe consiste à créer plusieurs versions de réponses attendues avec l'IA générative afin d'améliorer l'évaluation d'une réponse étudiant. Mais en ce qui concerne la présente étude, la performance EQM demeure à l'avantage de simsemlis.

8 Conclusion

Nous proposons une méthode qui offre l'avantage d'une performance près de l'état de l'art en termes de précision. Elle est aussi très peu gourmande en termes de ressources de calculs comparativement aux autres approches de l'état de l'art qui nécessitent un affinement, et elle est d'une grande simplicité à implémenter. Cependant, elle nécessite une correction manuelle d'une quinzaine de réponses par question.

Cependant, une analyse de la distribution des scores par question suggère que l'étape de correction manuelle pour calibrer la correction ou pour affiner un modèle est peut-être incontournable.

Mais que la correction manuelle d'une partie des réponses soit incontournable ou non, la responsabilité d'avoir une correction fiable repose toujours dans les mains de l'invidu qui corrige. Pour cette raison, une prochaine étape de cette étude encore exploratoire est de déterminer dans quelle mesure l'approche est en mesure d'identifier les évaluations automatiques qui sont fiables de celles qui nécessiteraient une vérification.

Références

1. Agarwal, R., Khurana, V., Grover, K., Mohania, M., Goyal, V. : Multi-relational graph transformer for automatic short answer grading. In : Carpuat, M., de Marneffe, M.C., Meza Ruiz, I.V. (eds.) Proceedings of the 2022 Conference of the

North American Chapter of the Association for Computational Linguistics : Human Language Technologies. pp. 2001–2012. Association for Computational Linguistics, Seattle, United States (Jul 2022). https://doi.org/10.18653/v1/2022.naacl-main.146, https://aclanthology.org/2022.naacl-main.146/

2. Aggarwal, D., Bhattacharyya, P., Raman, B. : "I understand why I got this grade" : Automatic short answer grading with feedback (2024), https://arxiv.org/abs/2407.12818
3. Amur, Z.H., Kwang Hooi, Y., Bhanbhro, H., Dahri, K., Soomro, G.M. : Short-text semantic similarity (stss) : Techniques, challenges and future perspectives. Applied Sciences **13**(6), 3911 (2023)
4. Banarescu, L., Bonial, C., Cai, S., Georgescu, M., Griffitt, K., Hermjakob, U., Knight, K., Koehn, P., Palmer, M., Schneider, N. : Abstract meaning representation for sembanking. In : Proceedings of the 7th linguistic annotation workshop and interoperability with discourse. pp. 178–186 (2013)
5. Carpenter, D., Min, W., Lee, S., Ozogul, G., Zheng, X., Lester, J. : Assessing student explanations with large language models using fine-tuning and few-shot learning. In : Proceedings of the 19th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2024). pp. 403–413 (2024)
6. Chang, L.H., Ginter, F. : Automatic short answer grading for Finnish with ChatGPT. In : Proceedings of the AAAI Conference on Artificial Intelligence. vol. 38, pp. 23173–23181 (2024)
7. Condor, A., Litster, M., Pardos, Z. : Automatic short answer grading with sbert on out-of-sample questions. International Educational Data Mining Society (2021), https://educationaldatamining.org/EDM2021/virtual/static/pdf/EDM21_paper_149.pdf
8. Del Gobbo, E., Guarino, A., Cafarelli, B., Grilli, L. : GradeAid : a framework for automatic short answers grading in educational contexts—design, implementation and evaluation. Knowledge and Information Systems **65**(10), 4295–4334 (2023), https://rdcu.be/d5aDn
9. Divya, A., Haridas, V., Narayanan, J. : Automation of short answer grading techniques : Comparative study using deep learning techniques. In : 2023 Fifth International Conference on Electrical, Computer and Communication Technologies (ICECCT). pp. 1–7 (2023). https://doi.org/10.1109/ICECCT56650.2023.10179759, https://ieeexplore.ieee.org/document/10179759
10. Firoozi, T., Bulut, O., Epp, C.D., Naeimabadi, A., Barbosa, D. : The effect of fine-tuned word embedding techniques on the accuracy of automated essay scoring systems using neural networks. Journal of Applied Testing Technology pp. 21–29 (2022)
11. Gaddipati, S.K., Nair, D., Plöger, P.G. : Comparative evaluation of pretrained transfer learning models on automatic short answer grading. arXiv preprint arXiv :2009.01303 (2020)
12. Garg, J., Papreja, J., Apurva, K., Jain, G. : Domain-specific hybrid BERT based system for automatic short answer grading. In : 2022 2nd International Conference on Intelligent Technologies (CONIT). pp. 1–6. IEEE (2022)
13. Ghavidel, H.A., Zouaq, A., Desmarais, M.C. : Using BERT and XLNET for the automatic short answer grading task. In : CSEDU (1). pp. 58–67 (2020)

14. Grévisse, C. : Llm-based automatic short answer grading in undergraduate medical education. BMC Medical Education **24**(1), 1060 (2024)
15. Mohler, M., Bunescu, R., Mihalcea, R. : Learning to grade short answer questions using semantic similarity measures and dependency graph alignments. In : Lin, D., Matsumoto, Y., Mihalcea, R. (eds.) Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics : Human Language Technologies. pp. 752–762. Association for Computational Linguistics, Portland, Oregon, USA (Jun 2011), https://aclanthology.org/P11-1076/
16. Mohler, M., Mihalcea, R. : Text-to-text semantic similarity for automatic short answer grading. In : Lascarides, A., Gardent, C., Nivre, J. (eds.) Proceedings of the 12th Conference of the European Chapter of the ACL (EACL 2009). pp. 567–575. Association for Computational Linguistics, Athens, Greece (Mar 2009), https://aclanthology.org/E09-1065/
17. Ouahrani, L., Bennouar, D. : Paraphrase generation and supervised learning for improved automatic short answer grading. International Journal of Artificial Intelligence in Education pp. 1–44 (2024)
18. Sultan, M.A., Salazar, C., Sumner, T. : Fast and easy short answer grading with high accuracy. In : Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics : Human Language Technologies. pp. 1070–1075 (2016)
19. Tobler, S. : Smart grading : A generative ai-based tool for knowledge-grounded answer evaluation in educational assessments. MethodsX **12**, 102531 (2024)
20. Tulu, C.N., Ozkaya, O., Orhan, U. : Automatic short answer grading with semspace sense vectors and malstm. IEEE Access **9**, 19270–19280 (2021)
21. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I. : Attention is all you need. Advances in neural information processing systems **30** (2017)
22. Yoon, S.Y. : Short answer grading using one-shot prompting and text similarity scoring model (2023), https://arxiv.org/abs/2305.18638
23. Zhu, X., Wu, H., Zhang, L. : Automatic short-answer grading via BERT-based deep neural networks. IEEE Transactions on Learning Technologies **15**(3), 364–375 (2022)

Simulations scientifiques pour les travaux pratiques : choix de conception et affordances pédagogiques

Cédric d'Ham[1], Sébastien Rebaudo[1], Rémi Colin de Verdière[2]

[1] LIG-MeTAH, Université Grenoble Alpes, CNRS, Grenoble INP
[2] Faculté des Sciences, Université Grenoble Alpes
cedric.dham@univ-grenoble-alpes.fr

Résumé. Les travaux pratiques (TP) sont essentiels pour l'apprentissage des sciences. Les TP simulés, utilisant des simulations informatiques à la place d'expérimentations physiques, sont de plus en plus utilisés en combinaison avec les TP en laboratoire pour en améliorer l'efficacité pédagogique. Nous analysons, sous l'angle de la transposition informatique, les affordances pédagogiques d'un TP simulé prototypique avant de présenter une nouvelle approche des TP simulés. Dans celle-ci, la simulation est manipulée par l'intermédiaire d'un protocole expérimental élaboré par l'apprenant. Cette approche augmente la complexité de la tâche de l'apprenant, mais contribue à enrichir ses apprentissages. Une infrastructure logicielle a été développée pour produire des TP simulés basés sur notre nouvelle approche.

Mots-clés : enseignement des sciences, travaux pratiques, simulations, transposition informatique, affordances pédagogiques.

Abstract. Laboratories are essential for learning science. Simulated experiments, which use computer simulations instead of physical experiments, are increasingly combined with science labs to enhance pedagogical effectiveness. We analyze the pedagogical affordances of a prototypical simulated experiment through the lens of computational transposition before introducing a new approach to simulated experiments. In this approach, the simulation is manipulated via an experimental protocol designed by the learner. This approach increases the complexity of the learner's task but promotes a richer and more diverse learning experience. A software framework has been developed to produce simulated experiments based on our new approach.

Keywords: science laboratory, simulations, computational transposition, pedagogical affordances.

1 Introduction

Les travaux pratiques (TP) constituent un élément central de l'enseignement des sciences [19]. Au cours de ces séances, les étudiants réalisent des expérimentations,

c'est-à-dire qu'ils « installent et manipulent des équipements dans le but de collecter des données » [10], données ensuite utilisées pour établir des résultats expérimentaux. Bien que l'efficacité des TP en matière d'apprentissage ait été régulièrement questionnée [2, 27], les enseignants restent attachés à cette modalité qu'ils considèrent comme centrale dans la formation scientifique [2].

L'idée de remplacer les expérimentations par des simulations est ancienne [8]. Dans les « TP simulés », les données ne sont pas mesurées par des instruments physiques, mais générées par un modèle. Une expérimentation consiste alors à choisir la valeur de paramètres, puis à lancer une simulation qui calcule des données considérées comme expérimentales. Une représentation du phénomène étudié est souvent affichée pour permettre à l'apprenant de faire le lien entre les données simulées et le phénomène (ou tout du moins sa représentation à l'écran). Comparées aux TP en laboratoire, les simulations offrent de nombreux avantages. Elles réduisent les coûts liés aux équipements, aux consommables et à l'encadrement. Elles rendent accessibles des expérimentations trop dangereuses, coûteuses ou complexes à mettre en œuvre. Enfin, il a été proposé que les simulations peuvent favoriser les apprentissages pour trois raisons principales [13] : (i) l'absence de contraintes matérielles permet de multiplier les expérimentations dans un temps réduit ; (ii) les données simulées, dépourvues d'erreurs expérimentales, sont plus faciles à analyser ; (iii) les visualisations peuvent être enrichies par la représentation de phénomènes inaccessibles à l'observation directe [24].

En raison des avantages attribués aux TP simulés, de nombreuses études ont comparé leur efficacité à celle des TP en laboratoire. Une revue de 2012 [26] rapporte un bénéfice clair en termes d'apprentissage de connaissances conceptuelles en faveur des TP simulés, lorsqu'ils remplacent des enseignements traditionnels ou des TP en laboratoire. Toutefois, les auteurs soulignent que la plupart des études présentent des biais en ne tenant pas compte des effets liés aux interventions des enseignants, aux documents de support et à la façon dont sont conçues les simulations. Une autre revue de 2013 [13] ne relève pas de différence significative entre les deux modalités et conclut à leur complémentarité, selon les phénomènes étudiés et les objectifs pédagogiques visés. Ces deux revues mettent en avant l'intérêt de combiner TP simulés et TP en laboratoire, que ce soit pour préparer les expérimentations en laboratoire ou pour leur apporter un éclairage conceptuel. Depuis, de nombreuses recherches ont cherché à évaluer les effets de cette combinaison. Une revue [31] et une méta-analyse [29] récentes montrent un effet positif sur les apprentissages de connaissances conceptuelles lorsque les deux modalités sont associées, par rapport à l'utilisation d'une seule modalité. L'ordre de présentation (simulation ou laboratoire en premier) ne semble pas influencer les résultats. Wang *et al.* [29] suggèrent que l'enjeu réside moins dans la séquence que dans la capacité de chaque modalité à fournir les informations les plus utiles selon le contexte.

Dans notre contribution, nous considérons l'ensemble des objectifs pédagogiques des TP, en élargissant notre analyse au-delà des seuls apprentissages de connaissances conceptuelles habituellement évalués. Nous examinons dans quelle mesure les simulations classiques soutiennent ces divers objectifs, et proposons un nouveau type de TP simulés visant à étendre le champ de leurs affordances pédagogiques.

2 Objectifs pédagogiques des Travaux Pratiques en sciences

Nous avons extrait de la littérature les objectifs pédagogiques les plus fréquemment associés aux TP de sciences, au niveau lycée ou licence. Selon les études, les objectifs identifiés proviennent d'enquêtes auprès d'enseignants [1, 3, 16, 20] ou d'étudiants [16], de l'analyse de fascicules de TP [28] ou d'avis d'experts [19]. Ils sont listés ci-dessous, du moins au plus spécifique aux TP de sciences, certains apprentissages ne pouvant être abordés que dans ce cadre [30].

1. *Développer des compétences transversales*, telles que la collaboration, la gestion du temps ou la rigueur.
2. *Renforcer la motivation pour l'apprentissage des sciences,* un objectif souvent associé aux TP bien qu'il ne concerne pas directement un apprentissage disciplinaire.
3. *Mobiliser des concepts et modèles scientifiques*, qui constituent le cœur des connaissances conceptuelles visées en sciences. Ces apprentissages ne sont pas spécifiques aux TP puisqu'ils sont aussi l'objectif des cours et travaux dirigés. La littérature montre que, pendant la phase de manipulation, les étudiants se focalisent souvent sur la réussite technique au détriment de la compréhension [21, 27], ce qui limite l'atteinte de cet objectif, ou le reporte à la phase de rédaction.
4. *Identifier et nommer des phénomènes scientifiques*, en s'appuyant sur la perception directe d'objets ou d'événements caractéristiques d'un champ disciplinaire, par exemple un précipité, une colonie bactérienne, un laser, etc.
5. *Reconnaître et manipuler le matériel de laboratoire,* en développant des compétences techniques de laboratoire, notamment la gestuelle et les pratiques garantissant la qualité des données produites et la sécurité des individus.
6. *Connaître les méthodes expérimentales courantes,* telles que la modification d'une seule variable à chaque nouvelle expérimentation [14], l'utilisation de contrôles positifs et négatifs, d'étalons ou de références [4], ou encore le calcul d'incertitudes pour comparer des résultats. Ces méthodes, bien que fondamentales, sont rarement comprises ou même conscientisées par les étudiants lorsqu'elles sont simplement exécutées en suivant des consignes prescriptives fournies dans des fascicules qualifiés de « cookbook[1] » [2, 27].
7. *Comprendre en quoi l'application d'une démarche articulant théorie et expérimentation fonde la validité des savoirs scientifiques*, dans une perspective épistémologique qui se manifeste à la fois dans la conduite d'une démarche expérimentale et dans sa mise en forme lors de la communication des résultats expérimentaux.

La liste ci-dessus montre que l'intérêt pédagogique des TP dépasse la seule mobilisation de connaissances conceptuelles. Ce constat questionne le choix des revues mentionnées en introduction de restreindre la comparaison entre TP simulés et TP en laboratoire à l'évaluation de cet unique type d'apprentissage.

[1] À la manière d'une recette de cuisine, où les actions à réaliser sont listées chronologiquement.

3 Affordances pédagogiques d'une simulation prototypique

Les simulations peuvent-elles couvrir l'ensemble des objectifs pédagogiques des TP ? Rutten *et al.* [26] donnent une première réponse en indiquant que « l'acquisition de compétences de laboratoire constitue souvent un objectif en soi, qui ne peut pas être entièrement remplacé par des simulations ». Dans cette partie, nous proposons d'analyser une simulation prototypique afin d'identifier ses affordances pédagogiques, c'est-à-dire sa capacité à servir de support à divers objectifs pédagogiques [22]. Cette analyse peut être réalisée sous l'angle de la transposition informatique[2] [5, 6] qui décrit les transformations subies par un savoir lorsqu'il est intégré dans un EIAH. L'analyse de la transposition informatique nécessite d'expliciter comment les objets de savoir (les concepts et relations de la discipline) sont codés et réifiés à l'interface du logiciel pour être visualisés et manipulés par les apprenants. Les représentations retenues et les contraintes de manipulation déterminent les affordances pédagogiques du dispositif, et donc les apprentissages que la simulation peut favoriser ou limiter.

Pour sélectionner un exemple prototypique de simulation, nous avons retenu un thème classique de TP : le dosage spectrophotométrique. Sur PhET [25], plateforme de référence proposant 171 simulations pédagogiques, une recherche avec le mot-clé « spectrophotometer » renvoie à la simulation « Beer's Law Lab » (fig. 1). Celle-ci est accompagnée d'une fiche de TP intitulée « Visible Spectrophotometry of Nickel (II) Chloride », fournissant les consignes pour réaliser un dosage spectrophotométrique.

Commençons par identifier les principaux objets de savoir accessibles (données d'entrée de la simulation) et leur mode de manipulation :

- Longueur d'onde de mesure (λ) : saisie au clavier ou réglage par curseur.
- Largeur de la cuve de mesure (l) : réglage par curseur.
- Nature de la solution dans la cuve de mesure (S) : choix parmi 8 solutions via un menu déroulant.
- Concentration de la solution (C) : réglage par curseur.

La simulation représente également trois objets de savoir calculés à partir de ces données, non manipulables directement :

- Couleur de la solution et intensité (phénomènes – dépendent de S et C).
- Absorbance (A) ou transmittance (mesures – dépendent de S, l, λ et C).
- Couleur du rayon lumineux et intensité avant et après la cuve (phénomènes – dépendent de λ et A).

La simulation ne comporte pas de scénarisation pédagogique. C'est la fiche de consigne qui gère cet aspect. Le scénario, très directif, indique à l'apprenant les valeurs à entrer, les données à collecter et leur mode de traitement : il s'agit d'un TP de type « cookbook ».

À partir de la description de la simulation en termes d'objets de savoir, nous pouvons évaluer ses affordances pédagogiques au regard des objectifs disciplinaires des TP (objectifs 3 à 7). Les objectifs 1 (compétences transversales) et 2 (motivation) ne relèvent

[2] Le mot « transposition » est utilisé avec le sens qu'il a dans l'expression « transposition didactique ».

pas directement de la transposition informatique, car ils dépendent davantage du scénario pédagogique que de la manipulation des objets de savoir. Notons que pour la motivation (objectif 2), il est établi que les TP simulés ont un impact moins favorable que les TP en laboratoire [7].

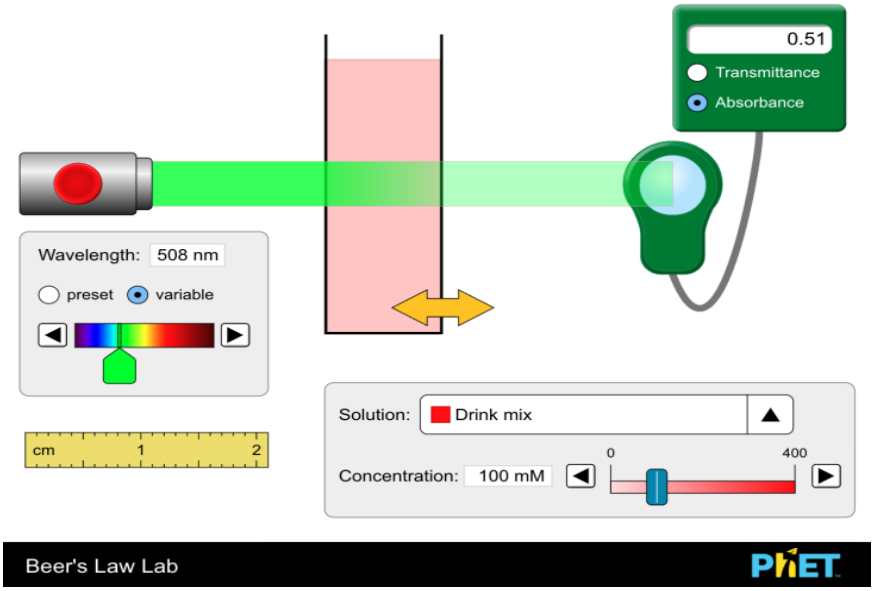


Fig. 1. Interface de la simulation de spectrophotomètre proposée par PhET.

Objectif 3 – Mobiliser des concepts et modèles. À partir des consignes de la fiche de TP et des manipulation possibles, la simulation engage les apprenants à mobiliser les concepts d'absorbance A et de spectre A = f(λ), ainsi que le modèle de Beer-Lambert $A = \varepsilon_{S\lambda}.l.C$. Des informations de réalité augmentée sont affichées aux apprenants : visualisation du rayon lumineux dont la couleur dépend de la longueur d'onde et dont l'intensité dépend de l'absorbance. Ces informations ont pour fonction de faciliter la compréhension du concept d'absorbance et des paramètres d'influence dans le modèle de Beer-Lambert.

Objectifs 4 et 5 – Identifier et nommer des phénomènes ; reconnaître et manipuler le matériel. La simulation ne soutient pas ces objectifs et risque même d'induire des conceptions erronées chez l'apprenant par les représentations du rayon lumineux visible et de la cuve de mesure non translucide.

Objectif 6 – Connaître les méthodes expérimentales courantes. Dans le contexte présenté, cet objectif n'est pas travaillé par les apprenants en raison des limites de la simulation et du scénario. La simulation ne requiert pas certaines actions essentielles, telles que l'enregistrement d'une référence. Les mesures d'absorbance, exemptes d'erreur, permettent de tracer une courbe étalon à partir de seulement deux points. Ces simplifications constituent des erreurs méthodologiques dans un environnement de laboratoire

réel et nuisent à la compréhension des exigences propres aux expérimentations. Au niveau de la fiche de consigne, la procédure, présentée sous forme prescriptive, relève d'une approche « cookbook », peu favorable à la compréhension des enjeux et des justifications des méthodes expérimentales [2, 27].

Objectif 7 – Comprendre en quoi l'articulation théorie/expérimentation fonde la validité des savoirs. Dans la mesure où la simulation remplace l'expérimentation par un modèle théorique, aucune réelle articulation entre théorie et pratique n'est engagée. Cet objectif ne peut donc être travaillé.

L'analyse d'une simulation classique montre qu'elle ne soutient que l'apprentissage des concepts et modèles (objectif 3). Cela explique sans doute pourquoi les études comparant simulations et TP en laboratoire se concentrent sur cet objectif. Par ailleurs, des recherches ayant montré que les connaissances conceptuelles sont généralement peu mobilisées par les apprenants lors des TP en laboratoire [21, 27], les simulations peuvent renforcer ce type d'apprentissage et donc avoir un effet synergique lors de leur combinaison avec ces TP [29, 31]. Cependant, nous avons identifié que la simulation analysée peut avoir des effets négatifs sur d'autres objectifs (notamment 4, 5 et 6), en sursimplifiant la situation expérimentale. Comme le notent certains auteurs, les apprenants peuvent alors construire une représentation trop simplifiée de ce qu'est une investigation expérimentale [9, 10]. Il apparaît évident qu'un TP simulé ne peut pas remplacer un TP en laboratoire.

4 Proposition d'un nouveau type de TP simulés

Dans la section précédente, nous avons analysé une simulation qualifiée de prototypique. Toutefois, les simulations ne sont pas toutes équivalentes et, comme certains auteurs le formulent, « l'espace pour la conception des simulations est très vaste » [26]. Dans notre équipe, nous développons un nouveau type de TP simulés visant à élargir le spectre des affordances pédagogiques de ces environnements.

La figure 2 illustre une partie de l'interface d'un TP simulé fondé sur notre modèle et portant sur le même thème que précédemment : le dosage spectrophotométrique. Nous analysons ici, sous l'angle de la transposition informatique, les principaux objets de savoir accessibles et leur mode de manipulation. C'est à ce niveau que se situe la principale différence avec la simulation PhET. Dans la partie supérieure de la capture, sont visibles les étapes 3 et 4 d'un protocole expérimental constitué d'une succession d'actions [17]. Ce protocole est construit par les apprenants en ajoutant et paramétrant ces actions à partir d'actions prédéfinies, à la manière du logiciel Scratch pour l'enseignement de l'informatique. Les paramètres des actions sont saisis au clavier (valeurs) ou sélectionnés via des menus (matériel). Par exemple, l'action paramétrée « Conditionner un(e) cuve en quartz… » est en cours de modification dans l'étape 4. Les actions du protocole et leurs paramètres sont les données d'entrée de la simulation.

Concernant les données calculées par la simulation, les seuls objets de savoir affichés sont les résultats fournis par le modèle. Par exemple, la partie inférieure de la figure 2 montre un spectre généré en réponse à une mesure spécifiée dans l'étape 2 du protocole

(non présentée dans la figure). Dans le TP « dosage spectrophotométrique », l'objet de savoir « dispositif expérimental » n'est pas représenté, bien qu'il puisse l'être dans d'autres TP simulés conçus selon notre modèle.

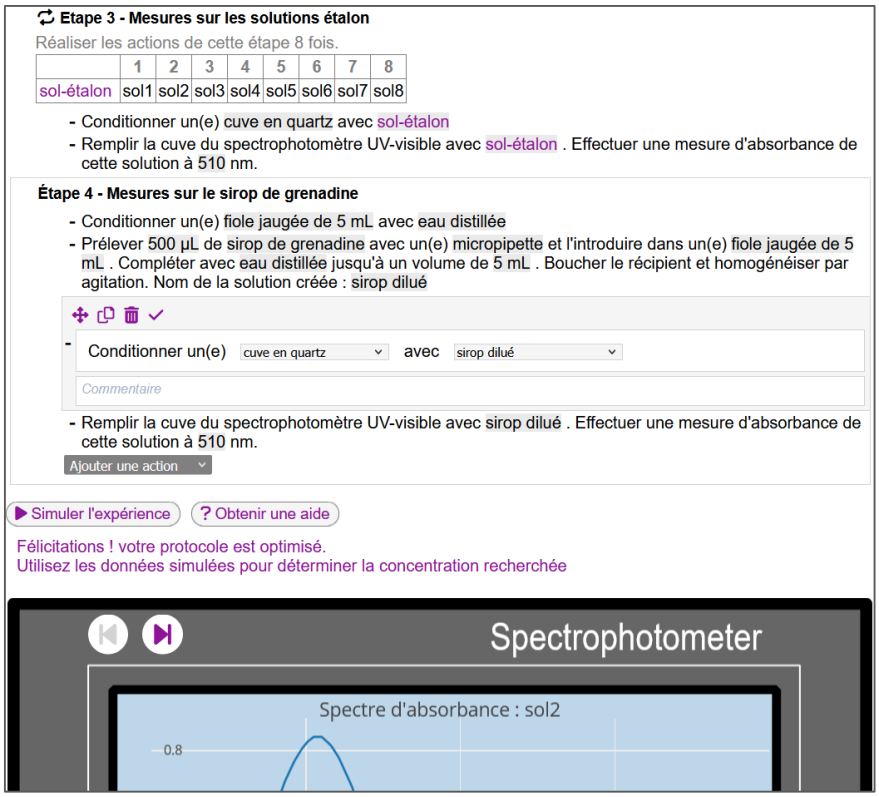


Fig. 2. Extrait de l'interface d'un TP simulé conçu selon notre modèle.

Contrairement aux simulations classiques, notre modèle intègre systématiquement un scénario pédagogique : à partir d'une question initiale, l'apprenant élabore un protocole expérimental, génère des données simulées, puis les exploite pour répondre au problème posé. L'élaboration d'un protocole pour définir les données d'entrée de la simulation rend le TP simulé nettement plus proche d'une activité de laboratoire que les simulations conventionnelles.

Evaluons les apports de notre TP simulé en termes d'affordances pédagogiques comparativement à la simulation PhET :

Objectif 3 – Mobiliser des concepts et modèles. Les mêmes concepts et modèles sont manipulés dans les deux situations : absorbance, spectres et loi de Beer-Lambert. Si la simulation PhET apporte une plus-value informationnelle via la réalité augmentée,

notre TP exige des apprenants qu'ils comprennent et mobilisent activement ces notions pour concevoir leur expérimentation.

Objectif 4 – Identifier et nommer des phénomènes. Aucun des deux dispositifs ne permet d'atteindre cet objectif.

Objectif 5 – Reconnaître et manipuler le matériel. L'aspect perceptivo-gestuel de la manipulation du matériel n'est pas accessible dans une simulation. Toutefois, notre modèle engage l'apprenant à choisir lui-même le matériel, nécessitant qu'il en comprenne la fonction.

Objectif 6 – Connaître les méthodes expérimentales courantes. Il s'agit du principal apport de notre TP simulé. L'élaboration d'un protocole par l'apprenant favorise l'acquisition de connaissances spécifiques à la pratique expérimentale et notamment au niveau des méthodes [15]. Cette activité, centrale dans notre modèle, n'est pas demandée dans le TP PhET, et rarement dans les TP en laboratoire car difficile à encadrer.

Objectif 7 – Comprendre en quoi l'articulation théorie/expérimentation fonde la validité des savoirs. Malgré la réflexion pratique requise pour élaborer un protocole, les résultats expérimentaux étant simulés, cet objectif reste difficilement atteignable.

En conclusion de notre analyse, une simulation classique permet de faire mobiliser, voire découvrir des concepts et modèles aux apprenants (objectif 3). Notre modèle permet lui aussi des apprentissages de connaissances conceptuelles mais également des apprentissages méthodologiques en termes de méthodes expérimentales (objectif 6) et de fonctions des matériels de laboratoire (une partie de l'objectif 5).

L'élargissement des affordances pédagogiques dans notre modèle repose sur l'activité de conception d'expérience confiée aux apprenants. Cette tâche, complexe, nécessite un étayage adapté [11, 18] afin que les apprenants produisent des protocoles suffisamment élaborés pour être simulables. Un premier étayage est assuré par la structuration du protocole, fondée sur des actions paramétrées. Un second est proposé via le bouton « Obtenir une aide » (fig. 2) : sur demande, le protocole est évalué par un système de contraintes [23] qui renvoie une rétroaction en fonction de l'historique de travail de l'apprenant.

Notre modèle de TP simulés est plus complexe à développer qu'une simulation classique. Pour faciliter ce développement, nous avons mis en place un framework logiciel qui réutilise les outils de la plateforme LabNBook [12], notamment ceux dédiés à l'élaboration de protocoles et au traitement de données. Des outils-auteur ont été intégrés dans ce framework pour que le concepteur puisse définir la structure du protocole, la spécification des contraintes à appliquer et les messages de rétroaction à destination de l'apprenant. Contraintes et simulations sont spécifiquement implémentées pour chaque TP. Deux TP ont été développés sur la base de ce framework (titrage acide-base et dosage spectrophotométrique) et ont été utilisés par 540 étudiants de licence 1 en préparation de séances de laboratoire où l'expérimentation est à concevoir. Trois autres TP simulés sont en cours de production (évaluation bactériologique d'une eau, dilution, dissolution).

5 Conclusion

Les travaux pratiques occupent une place centrale dans l'enseignement des sciences, en particulier pour développer des compétences qui dépassent la simple acquisition de connaissances conceptuelles. À partir d'une revue des objectifs pédagogiques usuellement associés aux TP, nous avons montré, par une analyse fondée sur la transposition informatique, que les simulations classiques ne permettent de soutenir qu'un nombre restreint de ces objectifs et en particulier celui lié aux apprentissages de connaissances conceptuelles.

Pour répondre à cette limite, nous avons proposé un nouveau modèle de TP simulés, centré sur la conception d'expérience par les apprenants. Ce modèle repose sur l'environnement numérique LabNBook, et permet d'étendre les affordances pédagogiques des simulations à des objectifs liés aux méthodes expérimentales et à la compréhension des fonctions du matériel. La structuration des protocoles sous forme d'actions paramétrées, associée à un système de rétroactions adaptatives, offre un étayage adapté à la complexité de la tâche. Un framework logiciel a été conçu pour permettre la production de tels TP simulés, en facilitant leur création par les enseignants et développeurs. Plusieurs TP ont déjà été réalisés et testés, et d'autres sont en cours de développement.

Cette dynamique ouvre de nombreuses perspectives de recherche : appropriation du framework par les enseignants, identification des caractéristiques d'un TP permettant de le transposer au format simulé, impact sur les apprentissages méthodologiques, ou encore optimisation de la stratégie de rétroaction pour soutenir la réussite et les apprentissages. Certaines de ces questions sont actuellement explorées au sein de notre équipe, tandis que d'autres restent à ouvrir pour de futurs travaux.

References

1. Abraham, M. R.: What can be learned from laboratory activities? Revisiting 32 years of research. *Journal of Chemical Education*, 88(8), 1020-1025 (2011).
2. Abrahams, I., Millar, R.: Does practical work really work? A study of the effectiveness of practical work as a teaching and learning method in school science. *International Journal of Science Education*, 30(14), 1945-1969 (2008).
3. Agustian, H. Y., Pedersen, M. I., Finne, L. T., Jørgensen, J. T., Nielsen, J. A., & Gammelgaard, B.: Danish university faculty perspectives on student learning outcomes in the teaching laboratories of a pharmaceutical sciences education. *Journal of Chemical Education*, 99(11), 3633-3643 (2022).
4. Baker, L. M., Dunbar, K.: Experimental design heuristics for scientific discovery: The use of "baseline" and "known standard" controls. *International Journal of Human-Computer Studies* 53(3), 335-349 (2000).
5. Balacheff, N.: La transposition informatique, un nouveau problème pour la didactique. In M. Artigue, R. Gras, C. Laborde, P. Tavignot (eds.) Vingt ans de didactique des mathématiques en France, pp. 364-370. La Pensée Sauvage, Grenoble (1993).
6. Balacheff, N., Kaput, J. J.: Computer-based learning environment in mathematics. In A. Bischop (ed.), International Handbook of Mathematics Education, pp. 469-501. Kluwer Academic publisher, Dordrecht (1997).

7. Corter, J. E., Esche, S. K., Chassapis, C., Ma, J., Nickerson, J. V.: Process and learning outcomes from remotely-operated, simulated, and hands-on student laboratories. *Computers & Education*, 57(3), 2054-2067 (2011).
8. Boblick, J. M.: The use of computer simulations in the teaching of high school physics. *Science Education* 54(1), 77-81 (1970).
9. Chen, S., Chang, W.-H., Lai, C.-H., Tsai, C.-Y.: A Comparison of Students' Approaches to Inquiry, Conceptual Learning, and Attitudes in Simulation-Based and Microcomputer-Based Laboratories. *Science Education* 98(5), 905-935 (2014).
10. Crippen, K. J., Archambault, L. M., Kern, C. L.: The nature of laboratory learning experiences in secondary science online. *Research in Science Education* 43(3), 1029-1050 (2013).
11. d'Ham C., Girault I., Berthet A.: Modèles et étayages pour l'élaboration de protocoles par les élèves : cas des titrages acide-base. *Recherche en Didactique des Sciences et des Technologies*, 19, 165-186 (2019).
12. d'Ham C., Wajeman C., Girault I., Marzin-Janvier P.: LabNBook, plateforme numérique support des pédagogies actives et collaboratives en sciences expérimentales. In : Actes de la 9ème Conférence sur les Environnements Informatiques pour l'Apprentissage Humain, Paris, pp. 49-60 (2019).
13. de Jong, T., Linn, M. C., Zacharia, Z. C.: Physical and virtual laboratories in science and engineering education. *Science* 340, 305-308 (2013).
14. Efstathiou, C., Hovardas, T., Xenofontos, N. A., Zacharia, Z. C., deJong, T., Anjewierden, A., van Riesen, S. A. N.: Providing guidance in virtual lab experimentation: The case of an experiment design tool. *Educational Technology Research and Development* 66(3), 767-791 (2018).
15. Etkina, E., Karelina, A., Ruibal-Villasenor, M., Rosengrant, D., Jordan, R., Hmelo-Silver, C. E.: Design and Reflection Help Students Develop Scientific Abilities: Learning in Introductory Physics Laboratories. *Journal of the Learning Sciences* 19(1), 54-98 (2010).
16. George-Williams, S. R., Ziebell, A. L., Kitson, R. R. A., Coppo, P., Thompson, C. D., Overton, T. L.: 'What do you think the aims of doing a practical chemistry course are?' A comparison of the views of students and teaching staff across three universities. *Chemistry Education Research and Practice* 19(2), 463-473 (2018).
17. Girault, I., d'Ham, C., Ney, M., Sanchez, E., Wajeman, C.: Characterizing the experimental procedure in science laboratories: a preliminary step toward students experimental design. *International Journal of Science Education* 34 (6), 825-854 (2012).
18. Girault, I., d'Ham, C.: Scaffolding a complex task of experimental design in chemistry with a computer environment. *Journal of Science Education and Technology* 23 (4), 514-526 (2014).
19. Hofstein, A., Lunetta, V. N.: The laboratory in science education: Foundations for the twenty-first century. *Science Education*, *88*(1), 28-54 (2004).
20. Josephsen, J.: Experimental training for chemistry students: does experimental experience from the general sciences contribute? *Chemistry Education Research and Practice*, 4(2), 205-218 (2003).
21. Kluge, A.: Combining Laboratory Experiments with Digital Tools to Do Scientific Inquiry. *International Journal of Science Education*, 36(13) 2157-2179 (2014).
22. Nagashima, T., Yang, K., Bartel, A., Silla, E., Vest, N., Alibali, M., & Aleven, V.: Pedagogical affordance analysis: Leveraging teachers' pedagogical knowledge to elicit pedagogical affordances and constraints of instructional tools. In *The Interdisciplinarity of the Learning Sciences, 14th International Conference of the Learning Sciences (ICLS) 2020* (Vol. 3) (2020).

23. Ohlsson, S.: Constraint-Based Student Modeling. In: Greer, J.E., McCalla, G.I. (eds) *Student Modelling: The Key to Individualized Knowledge-Based Instruction*. NATO ASI Series, vol 125. Springer, Berlin, Heidelberg (1994).
24. Olympiou, G., Zacharias, Z., de Jong, T.: Making the invisible visible: Enhancing students' conceptual understanding by introducing representations of abstract objects in a simulation. *Instructional Science* 41(3), 575-596 (2012).
25. PhET Homepage, https://phet.colorado.edu/, dernier accès le 9/05/2025.
26. Rutten, N., van Joolingen, W. R., van der Veen, J. T.: The learning effects of computer simulations in science education. *Computers & Education* 58(1), 136-153 (2012).
27. Séré, M.-G., Beney, M.: Le fonctionnement intellectuel d'étudiants réalisant des expériences : observation de séances de travaux pratiques en premier cycle universitaire scientifique. *Didaskalia* 11, 75–102 (1997).
28. Tiberghien, A., Veillard, L., Le Maréchal, J.-F., Buty, C., Millar, R.: An analysis of labwork tasks used in science teaching at upper secondary school and university levels in several European countries. *Science Education* 85(5), 483-508 (2001).
29. Wang, X.-M., Yu, X.-H., Yu, D.-D. Hwang, G.-J., Lan, M.: Does combining real and virtual experiments improve learning achievement in physics? Evidence from a meta-analysis (2001–2021). *Educational Research Review* 46 (2025).
30. White, R.T.: The link between the laboratory and learning. *International Journal of Science Education*, 18, 761-774 (1996).
31. Wörner, S., Kuhn, J., Scheiter, K.: The best of two worlds: A systematic review on combining real and virtual experiments in science education. *Review of Educational Research* 92(6), 911-952 (2022).

Effet de la compétition-compétitivité vs coopération-coopérativité sur la recherche d'informations et l'apprentissage sur Internet

Cheyenne Dosso[1], Jean-Sébastien Vayre[2], Mohamed S. Benlamine[3], Tiffany Morisseau[4] et Christophe Heintz[5]

[1] cheyenne.dosso@univ-cotedazur.fr
[2] jean-sebastien.vayre@univ-cotedazur.fr
[3] mbenlamine@hct.ac.ae
[4] tiffany.morisseau@gmail.com
[5] christophe.heintz@gmail.com

Résumé. Comment la compétition-compétitivité et la coopération-coopérativité influencent-elles les activités de recherche d'informations sur Internet et les apprentissages qui en découlent ? À l'aide d'une expérimentation menée en laboratoire auprès de 47 participants, nous montrons que, contrairement à la compétition-compétitivité, la coopération-coopérativité favorise l'engagement des utilisateurs dans des stratégies de recherche d'informations d'exploitation. Nous montrons que ces stratégies encouragent le développement des connaissances factuelles et d'une humilité épistémique se traduisant par des attitudes de plus en plus tournées vers les positions dont les arguments sont les plus solides scientifiquement.

Mots-clés : recherche d'informations, vigilance épistémique, compétition, compétitivité, coopération, coopérativité.

Abstract. How do competition-competitiveness and cooperation-cooperativeness influence Internet information-seeking activities and the resulting learning outcomes? Through a laboratory experiment involving 47 participants, we demonstrate that, unlike competition-competitiveness, cooperation-cooperativeness fosters user engagement in exploitative information-seeking strategies. These strategies, in turn, promote the development of factual knowledge and epistemic humility, reflected in attitudes increasingly oriented toward positions supported by the most scientifically robust arguments.

Keywords: information seeking, epistemic vigilance, competition, competitiveness, cooperation, cooperativeness.

1 Introduction

À l'ère du numérique, la quantité d'informations disponibles en ligne a transformé la façon dont les individus recherchent et acquièrent des connaissances. Alors que les utilisateurs naviguent dans ce paysage informationnel complexe, deux dynamiques sociales essentielles – la compétition et la coopération – jouent un rôle important dans la définition de leurs comportements en ligne et de leurs apprentissages. Comprendre ce rôle est essentiel pour développer des stratégies efficaces qui améliorent les expériences d'apprentissage et favorisent les compétences de pensée critique chez les utilisateurs dans les environnements numériques.

Cet article explore l'interaction entre la compétition-compétitivité et la coopération-coopérativité dans les activités de recherche d'informations en ligne. Au travers d'une expérience contrôlée en laboratoire impliquant 47 participants, nous étudions comment ces dynamiques sociales influencent les stratégies employées lors de la recherche d'informations et affectent ensuite l'acquisition de connaissances factuelles.

2 Recherche d'informations et vigilance épistémique : effet de la compétition-compétitivité vs coopération-coopérativité

2.1 Recherche d'informations et vigilance épistémique

L'activité de recherche d'informations sur Internet peut être présentée à l'aide de trois processus de base : la planification, l'évaluation et le traitement de l'information (*cf.* Dosso et al., 2023 ; Rouet et Tricot, 1998 ; Sanchiz et al., 2020 ; Sharit et al., 2015 ; Tricot et Comtat, 2012).

Lors de la phase de planification, l'utilisateur commence par construire une représentation mentale de la recherche qu'il souhaite effectuer afin d'identifier les informations qui lui manquent au regard de ses connaissances antérieures. Il se construit ainsi une représentation de son but d'apprentissage de façon à le communiquer au moteur de recherche utilisé.

Durant la phase d'évaluation, l'utilisateur compare la représentation mentale qu'il se fait du but de sa recherche avec les contenus présentés dans les pages de résultats (i.e., les *Search Engine Results Pages* ; SERPs) : il explore ces contenus afin de leur conférer des valeurs de pertinence (*cf.* Golman & Loewenstein, 2018). Plus ces contenus sont proches de la représentation mentale que l'utilisateur a de son besoin d'information, plus il leur attribue des valeurs de pertinence élevées et plus les liens correspondant à ces contenus ont des chances d'être sélectionnés.

Dans la phase de traitement, il s'agit de comprendre et d'évaluer la pertinence des informations communiquées dans le document sélectionné au regard de l'objectif d'apprentissage que l'utilisateur s'est construit lors de la phase de planification.

Ces trois processus de base (i.e., planification, évaluation et traitement) sont itératifs et s'articulent selon des stratégies de recherche d'informations qui sont plus ou moins exploratoires (Dosso et al., 2023 ; 2024). D'une manière générale, les travaux antérieurs montrent que :

- les stratégies d'exploration conduisent les utilisateurs à centrer leur recherche d'informations sur la phase d'évaluation des SERPs et à développer une connaissance plus superficielle du sujet qui les intéressent, ce qui se traduit par l'acquisition de connaissances diverses mais générales (Collins-Thompson et al., 2016 ; Dosso et al., 2023 ; Hardy et al., 2019 ; Palani et al., 2021 ; Sanchiz et al., 2020) ;
- les stratégies d'exploitation conduisent les utilisateurs à centrer leur recherche d'informations sur la phase de traitement, c'est-à-dire sur l'activité de lecture de documents numériques, à développer une compréhension plus profonde du sujet qui les intéresse et à intégrer des informations plus spécifiques et ciblées dans leurs structures de connaissances (Dosso et al., 2023 ; Gadiraju et al., 2018 ; Hardy et al., 2019 ; Pardi et al., 2020 ; Salmeron & Llorens, 2018 ; Sanchiz et al., 2020).

Les utilisateurs peuvent être plus ou moins vigilants vis-à-vis des informations rencontrées lors de leurs navigations sur Internet (Tricot et al., 2016). La plausibilité d'une information est évaluée en fonction des connaissances de l'individu, c'est-à-dire de l'ensemble des hypothèses qui font partie de son contexte épistémique (Mercier & Sperber, 2017 ; Sperber et al., 2010). Ces hypothèses peuvent toutefois être révisées dans le sens où une information est d'autant mieux acceptée qu'elle provient d'une source considérée comme crédible et que son contenu est interprété comme cohérent (Braten et al., 2011 ; Brante & Stromso, 2018 ; Braten et al., 2017 ; Mason et al., 2010 ; Petty et Cacioppo, 1986 ; Stromso et al., 2013).

2.2 Effet de la compétition-compétitivité vs coopération-coopérativité sur le traitement de l'information et l'apprentissage

La théorie du traitement motivé de l'information (*cf.* Toma et al., 2012) soutient que le contexte relationnel à l'intérieur duquel se trouvent les utilisateurs influe sur la manière dont ils traitent l'information.

La compétition est une forme d'interdépendance négative qui entraîne une interaction d'opposition où chacun des individus cherche à entraver les efforts d'autrui (Deutsch, 1949 ; Johnson et Johnson, 1989).

Dans le domaine des apprentissages, il est établi que la relation de compétition et l'attitude qui y est associée – c'est-à-dire la compétitivité (*cf.* Xie et al., 2006 ; Lu et al., 2013) – favorisent le développement de buts de performance qui n'encouragent pas les utilisateurs à développer des stratégies de compréhension et de raisonnement cognitif et moral de hauts niveaux et qui les poussent à faire preuve de scepticisme vis-à-vis de la compétence et de la bienveillance d'autrui (Ames, 1992 ; Butera & Mugny, 1995 ; Dweck, 1986 ; Johnson & Johnson, 1985 ; Maehr, 1984 ; Nicholls, 1984 ; Skon, Johnson & Johnson, 1981 ; Toma, Gilles & Butera, 2011 ; Toma, Vasiljevic, Oberlé & Butera, 2012).

La coopération est une forme d'interdépendance positive qui entraîne une interaction de promotion où il s'agit d'encourager et de soutenir les efforts d'autrui (Deutsch, 1949 ; Johnson et Johnson, 1989).

Dans le domaine des apprentissages, la relation de coopération et l'attitude qui y est associée – c'est-à-dire la coopérativité (*cf.* Xie et al., 2006 ; Lu et al., 2013) – renforcent des stratégies tournées vers la maîtrise et la compréhension (i.e., les buts de maîtrise). Elle encourage l'acquisition de nouvelles connaissances et de meilleures capacités de raisonnement sur les problèmes que les individus rencontrent, y compris lorsque ces problèmes ont une dimension morale ou politique (Ames, 1992 ; Butera & Mugny, 1995 ; Dweck, 1986 ; Johnson & Johnson, 1985 ; Maehr, 1984 ; Nicholls, 1984 ; Skon, Johnson & Johnson, 1981 ; Toma, Gilles & Butera, 2011 ; Toma, Vasiljevic, Oberlé & Butera, 2012).

2.3 Hypothèses de recherche

Nous faisons l'hypothèse que la coopération-coopérativité conduit davantage les utilisateurs à développer des buts de maîtrise que la compétition-compétitivité. Ces buts de maîtrise se manifestent concrètement par :

a. un engagement plus important dans des stratégies de recherche d'informations d'exploitation ;
b. le développement d'un scepticisme moins important vis-à-vis de la crédibilité des sources et des contenus des informations communiquées lors de la recherche d'informations ;
c. le développement d'un plus grand nombre de connaissances factuelles ;
d. le développement d'une plus grande humilité épistémique qui se traduit par des attitudes de plus en plus tournées vers les positions dont les arguments sont les plus solides scientifiquement.

3 Méthodologie

3.1 Participants

L'étude que nous présentons est préliminaire et a été conduite auprès de 47 étudiants de l'Université Côte d'Azur. 26 participants ont été assignés au groupe « compétition » et 21 ont été assignés au groupe « coopération » de façon aléatoire mais en fonction de leur attitude vis-à-vis le consommation de protéine animale[1]. Les caractéristiques sociodémographiques de ces deux groupes sont présentées dans le tableau 1.

[1] Le déséquilibre d'effectif entre les deux groupes est lié au fait que les résultats que nous présentons sont préliminaires et qu'ils ont été produits à partir de l'effectif total de participants recrutés en juillet 2024.

Tableau 1. Présentation des caractéristiques sociodémographiques des participants

	Compétition	**Coopération**
Effectif	26	21
Âge moyen	28.8 (*SD* = 4.52)	23.8 (*SD* = 2.29)
Âge minimum	18	18
Âge maximum	42	25
Langue maternelle : français	23	20
Bilingue en français	3	1
Bac +1	9	10
Bac +2	7	6
Bac +3	10	5

Afin d'éviter les biais qui pourraient être associés à la composition des groupes « compétition » et « coopération », nous avons contrôlé l'ensemble des variables présentées dans le tableau 2.

Tableau 2. Présentation des moyennes et des écart-types correspondant aux variables de contrôle

	Compétition	**Coopération**
Nombre total d'années d'utilisation de l'internet	11.5 (*SD* = 3.23)	12.2 (*SD* = 3.37)
Auto-évaluation des connaissances en matière de recherche d'information (de 1 à 4)	2.92 (*SD* = .56)	2.86 (*SD* = .66)
Score total d'auto-efficacité en matière de recherche d'information (de 10 à 40)	32.2 (*SD* = 4.41)	34 (*SD* = 5.13)
Score total aux QCM préalables à l'étude en matière de recherche d'information	9.23 (*SD* = 3.44)	8.33 (*SD* = 2.8)

3.2 Protocole expérimental

Avant de réaliser le protocole expérimental, les participants ont répondu à un questionnaire en ligne permettant de recueillir leurs caractéristiques sociodémographiques, d'évaluer leurs attitudes vis-à-vis de la compétition et de la coopération (i.e., leurs compétitivité/coopérativité ; *cf.* Xie et al., 2006 ; Lu et al., 2013), leurs niveaux de connaissance et leurs attitudes vis-à-vis de la consommation de protéines animales. Nous avons choisi ce sujet car il est controversé scientifiquement et qu'il n'existe pas de solution simple aux problèmes sanitaires, écologiques, sociaux et économiques qui y sont associés.

Lors du protocole expérimental, les participants des groupes « compétition » et « coopération » ont dû réaliser en laboratoire, en synchrone (par groupe de 3 ou 4) et individuellement, des recherches sur Internet pendant 20 minutes sur le thème de la

consommation de protéines animales. Nous leur avons demandé d'effectuer ces recherches afin de préparer soit un débat compétitif (*cf.* le groupe « compétition ») soit une discussion coopérative (*cf.* le groupe « coopération ») réalisée par groupe de 3 ou 4. Le tableau 3 présente les instructions de ces deux tâches de recherche. À la suite de leurs recherches d'information, nous avons demandé aux participants de répondre aux mêmes questions qui leur étaient posées dans le questionnaire en ligne afin de réévaluer leurs niveaux de connaissance et leurs attitudes vis-à-vis de la consommation de protéines animales (*cf.* annexe 1).

Tableau 3. Présentation des instructions de recherche d'informations pour les groupes « coopération » et « compétition »

Coopération	Compétition
Vous devez maintenant vous engager dans une recherche d'informations sur Internet pendant 20 minutes afin de préparer une discussion portant sur le sujet de la consommation de protéines d'origine animale. Votre objectif est de coopérer avec les autres pour améliorer collectivement vos connaissances et trouver un consensus sur le sujet suivant : doit-on ou non consommer de la viande ?	Vous devez maintenant vous engager dans une recherche d'informations sur Internet pendant 20 minutes afin de préparer un débat portant sur le sujet de la consommation de protéines d'origine animale. Votre objectif personnel est de convaincre les autres de l'exactitude et du bien-fondé de votre point de vue sur le sujet suivant : doit-on ou non consommer de la viande ?
Pour vous aider, nous vous proposons d'imaginer la situation suivante : les autorités françaises réfléchissent à l'instauration d'une nouvelle loi visant à interdire définitivement à la population de consommer de la viande et qu'il soit plus compliqué de trouver des produits d'origine animale (lait, œufs...) dans les commerces. En tant que citoyen.ne, vous avez la possibilité de délibérer ensemble du bien-fondé de cette loi mais aussi de proposer ensemble des alternatives.	Pour vous aider, nous vous proposons d'imaginer la situation suivante : les autorités françaises ont décidé d'interdire définitivement à la population de consommer de la viande et il sera désormais plus compliqué de trouver des produits d'origine animale (lait, œufs...) dans les commerces. En tant que citoyen.ne, vous avez le devoir de faire entendre votre opinion vis-à-vis de cette loi et de l'exprimer face aux autres participants afin de défendre vos droits et intérêts.
En fonction du niveau de connaissances acquises par le groupe et de l'atteinte d'un consensus sur le sujet, tous les membres du groupe (vous y compris) pourront bénéficier d'une gratification financière de 10€, en plus des 20€ initialement prévus.	En fonction de vos performances et de vos capacités à convaincre, vous pourrez bénéficier d'une gratification financière de 10€, en plus des 20€ initialement prévus. Plus vous imposerez votre avis de manière convaincante lors du débat, plus vous serez susceptible de gagner de l'argent.
Pendant toute la durée de votre recherche d'informations, vous aurez le droit de prendre des notes afin de préparer la discussion.	Pendant toute la durée de votre recherche d'informations, vous aurez le droit de prendre des notes afin de préparer le débat.

3.3 Variables dépendantes

Nous avons utilisé 12 variables dépendantes afin de pouvoir tester nos hypothèses. Ces 12 variables sont présentées sommairement dans le tableau 4. La présentation détaillée de ces variables est exposée dans l'annexe 2.

Tableau 4. Présentation des variables dépendantes

Indicateurs de mesure des stratégies d'exploration-exploitation navigationnelle et thématique
Vd_1 : proportion d'exploration navigationnelle
Vd_2 : proportion de temps passé à l'exploration navigationnelle
Vd_3 : proportion d'exploration thématique
Vd_4 : proportion d'exploitation thématique
Vd_5 : proportion de re-exploitation thématique
Indicateurs de mesure de vigilance épistémique
Vd_6 : score moyen d'évaluation de l'expertise de la source
Vd_7 : score moyen d'évaluation de la bienveillance de la source
Indicateur de mesure de gain de connaissances factuelles
Vd_8 : gain de connaissances factuelles
Indicateurs de mesure de changement-renforcement d'attitude
Vd_9 : changement-renforcement de l'attitude à l'égard de la science et de la santé (SSA)
Vd_10 : changement-renforcement de l'attitude à l'égard de la souffrance animale et de la consommation de vêtement (VAHC)
Vd_11 : changement-renforcement de l'attitude à l'égard de la souffrance animale et de la violence dans les abattoirs (VASA)
Vd_12 : changement-renforcement de l'attitude à l'égard de la politique et de l'écologie (EP)

3.4 Variables indépendantes

Les deux variables indépendantes que nous avons mobilisées sont le contexte relationnel – c'est-à-dire les situations de compétition et de coopération – et l'attitude vis-à-vis de ce contexte – c'est-à-dire la compétitivité et la coopérativité. Afin de pouvoir évaluer cette attitude, nous avons utilisé l'échelle de personnalité compétitive et coopérative (CCPS ; *cf.* annexe 3 ; Xie et al., 2006 ; Lu et al., 2013). Nous avons testé la validité de cette échelle à l'aide de l'Alpha de Cronbach (*cf.* tableau 5).

Tableau 5. Alpha de Cronbach relatif à l'échelle de compétitivité-coopérativité

	Compétitivité	**Coopérativité**
Alpha de Cronbach	.86 (M = 5.36 ; *SD* = .73)	.83 (M = 3.91 ; *SD* = 1.07)

4 Résultats

4.1 Effet de la compétition-compétitivité vs coopération-coopérativité sur les stratégies d'exploration-exploitation navigationnelle et thématique

Comme l'indique le tableau 6, le contexte relationnel a un effet significatif sur les stratégies d'exploration-exploitation navigationnelle et thématique déployées par les participants. Les participants en situation de compétition ont consacré plus de temps à l'exploration navigationnelle (M = .40 ; SD = .25 ; p < .05) que ceux en situation de coopération (M = .25 ; SD = .24 ; p < .05). Comparativement aux participants en situation de coopération (M = .02 ; SD = .07 ; p = .05), les participants en situation de compétition (M = .08 ; SD = .14 ; p = .05) ont également plus tendance à réutiliser les requêtes précédemment générées au cours de leurs sessions de recherche. Le tableau 7 montre que l'attitude vis-à-vis de la compétition-coopération (i.e., la compétitivité et la coopérativité) n'a pas d'influence significative sur les stratégies d'exploration-exploitation navigationnelle et thématique développées par les participants.

Tableau 6. Mesures des effets de la compétition-coopération sur les stratégies d'exploration-exploitation navigationnelle et thématique

	Compétition	Coopération	Mesure
Vd_1	.56 (*SD* = .16)	.55 (*SD* = .16)	$t_{(45)} = .358$; $p = .722$
Vd_2	.40 (*SD* = .25)	.25 (*SD* = .24)	$t_{(43,4)} = -2.148$; $p = .037$
Vd_3	.74 (*SD* = .18)	.76 (*SD* = .19)	$t_{(41,9)} = -.418$; $p = .678$
Vd_4	.17 (*SD* = .17)	.21 (*SD* = .19)	$t_{(4,6)} = -.809$; $p = .423$
Vd_5	.08 (*SD* = .14)	.02 (*SD* = .07)	$t_{(37,7)} = 2.077$; $p = .045$

Tableau 7. Mesures des effets de la compétitivité-coopérativité sur les stratégies d'exploration-exploitation navigationnelle et thématique

	E	*SE*	*t*	p	x^2	*p*
Vd_1. Modèle linéaire généralisé 1						
Coopérativité (cognition)	-3.371e-05	6.400e-03	-.005	.9958		
Coopérativité (comportement)	9.174e-03	6.803e-03	1.349	.1851		
Coopérativité (affect)	-4.500e-03	9.921e-03	-.454	.6526	.064	.999
Compétitivité (cognition)	-1.566e-03	6.427e-03	-.244	.8087		
Compétitivité (comportement)	-6.480e-04	9.118e-03	-.071	.9437		
Compétitivité (affect)	1.998e-04	5.845e-03	.034	.9729		
Vd_1. Modèle linéaire généralisé 2						
Coopérativité	.002267	.002533	.895	.3755	.031	.998
Compétitivité	-.000988	.002231	-.443	.6601		
Vd_2. Modèle linéaire généralisé 1						
Coopérativité (cognition)	-.008208	.010329	-.795	.432		
Coopérativité (comportement)	.002882	.010980	.262	.794		
Coopérativité (affect)	-.006570	.016012	-.410	.684	.208	.999
Compétitivité (cognition)	-.008654	.010373	-.834	.409		
Compétitivité (comportement)	-.003551	.014717	-.241	.811		
Compétitivité (affect)	-.003049	.009434	-.323	.748		
Vd_2. Modèle linéaire généralisé 2						
Coopérativité	-.004034	.004051	-.996	.3248	.170	.982
Compétitivité	-.005333	.003569	-1.494	.1423		
Vd_3. Modèle linéaire généralisé 1						
Coopérativité (cognition)	.016609	.007121	2.333	.0248		
Coopérativité (comportement)	-.004054	.007569	-.536	.5952		
Coopérativité (affect)	-.009516	.011038	-.862	.3937	.224	.999
Compétitivité (cognition)	.010361	.007151	1.449	.1552		
Compétitivité (comportement)	-.002537	.010145	-.250	.8038		
Compétitivité (affect)	-.003269	.006504	-.503	.6180		
Vd_3. Modèle linéaire généralisé 2						
Coopérativité	.003600	.002951	1.220	.2290	.051	.996
Compétitivité	.001247	.002600	.480	.6338		
Vd_4. Modèle linéaire généralisé 1						
Coopérativité (cognition)	-.022729	.006166	-3.686	.000676		
Coopérativité (comportement)	.007163	.006555	1.093	.281025		
Coopérativité (affect)	.020551	.009559	2.150	.037655	.415	.999
Compétitivité (cognition)	-.009505	.006192	-1.535	.132662		
Compétitivité (comportement)	.006369	.008785	.725	.472675		
Compétitivité (affect)	.005375	.005632	.954	.345596		
Vd_4. Modèle linéaire généralisé 2						
Coopérativité	-.002220	.002826	-.786	.436	.036	.998
Compétitivité	.001347	.002490	.541	.591		
Vd_5. Modèle linéaire généralisé 1						
Coopérativité (cognition)	.0061199	.0045893	1.334	.190		
Coopérativité (comportement)	-.0031092	.0048784	-.637	.528		
Coopérativité (affect)	-.0110349	.0071141	-1.551	.129	.086	.999
Compétitivité (cognition)	-.0008553	.0046087	-.186	.854		
Compétitivité (comportement)	-.0038327	.0065386	-.586	.561		
Compétitivité (affect)	-.0021063	.0041916	-.503	.618		
Vd_5. Modèle linéaire généralisé 2						
Coopérativité	-.001380	.001869	-.738	.464	.035	.998
Compétitivité	-.002594	.001646	-1.576	.122		

4.2 Effet de la compétition-compétitivité vs coopération-coopérativité sur la vigilance épistémique

Les tableau 9 montrent que, contrairement au contexte relationnel (*cf.* tableau 8), l'attitude à l'égard de la compétition-coopération a un effet significatif sur la vigilance épistémique. Comparativement à la compétitivité qui n'a pas d'influence significative sur la vigilance épistémique, la coopérativité diminue le scepticisme des participants à l'égard de la bienveillance des sources qu'ils consultent (i.e., elle est évaluée plus positivement ; E = .163 ; t = 2.092 ; p < .05 ; χ^2 = 219.2 ; p < .001).

Tableau 8. Mesures des effets de la compétition-coopération sur la vigilance épistémique

	Compétition	Coopération	Mesure
Vd_6	26.5 (*SD* = 3.14)	24.5 (*SD* = 6.93)	$t_{(26.6)}$ = 1.221 ; p = .233
Vd_7	28.4 (*SD* = 2.73)	26.03 (*SD* = 7.11)	$t_{(24.8)}$ = 1.419 ; p = .168

Tableau 9. Mesures des effets de la compétitivité-coopérativité sur la vigilance épistémique

	E	*SE*	*t*	*p*	x^2	*p*
Vd_6. Modèle linéaire généralisé 1						
Coopérativité (cognition)	.06770	.20770	.326	.746	134.7	<.001
Coopérativité (comportement)	.18554	.22079	.840	.406		
Coopérativité (affect)	.14763	.32197	.459	.649		
Compétitivité (cognition)	.04067	.20858	.195	.846		
Compétitivité (comportement)	-.23663	.29592	-.800	.429		
Compétitivité (affect)	-.04177	.18970	-.220	.827		
Vd_6. Modèle linéaire généralisé 2						
Coopérativité	.13618	.08151	1.671	.10188	118.2	<.001
Compétitivité	-.06549	.07181	-.912	.36673		
Vd_7. Modèle linéaire généralisé 1						
Coopérativité (cognition)	.110822	.198423	.559	.57961	237.8	<.001
Coopérativité (comportement)	.164349	.210924	.779	.44046		
Coopérativité (affect)	.228808	.307584	.744	.46129		
Compétitivité (cognition)	-.002271	.199262	-.011	.99096		
Compétitivité (comportement)	-.315499	.282703	-1.116	.27108		
Compétitivité (affect)	-.080267	.181230	-.443	.66022		
Vd_7. Modèle linéaire généralisé 2						
Coopérativité	.16319	.07800	2.092	.04224	219.2	<.001
Compétitivité	-.11434	.06872	-1.664	.10325		

4.3 Effet de la compétition-compétitivité vs coopération-coopérativité sur le développement des connaissances factuelles

À l'inverse de l'attitude, le contexte relationnel influe significativement sur le développement des connaissances factuelles (*cf.* les tableaux 10 et 11). Les participants en situation de compétition ont un gain de connaissances factuelles presque dix fois plus faible (M = 0.23 ; SD = 3 ; p < .05) que ceux en situation de coopération (M = 2.1 ; SD = 2.72 ; p < .05).

Tableau 10. Mesures des effets de la compétition-coopération sur le développement de connaissances factuelles

	Compétition	Coopération	Mesure
Vd_8	.23 (*SD* = 3)	2.1 (*SD* = 2.72)	$t_{(45)} = -2.21$; $p = .032$

Tableau 11. Mesures des effets de la compétitivité-coopérativité sur le développement de connaissances factuelles[2]

	E	*SE*	*t*	*p*	x^2	*p*
Vd_8. Modèle linéaire généralisé 1						
Coopérativité (cognition)	.060097	.122289	.491	.626	25.02	<.001
Coopérativité (comportement)	-.070791	.129993	-.545	.589		
Coopérativité (affect)	.003899	.189565	.021	.984		
Compétitivité (cognition)	-.010554	.122806	-.086	.932		
Compétitivité (comportement)	.228041	.174231	1.309	.198		
Compétitivité (affect)	-.125323	.111692	-1.122	.269		
Vd_8. Modèle linéaire généralisé 2						
Coopérativité	.008699	.049086	.177	.860	1.13	.769
Compétitivité	.014379	.043245	.332	.741		

4.4 Effet de la compétition-compétitivité vs coopération-coopérativité sur le changement-renforcement d'attitude

Les tableaux 12 et 13 montrent un effet significatif du contexte relationnel et de l'attitude à l'égard de la compétition-coopération sur le changement-renforcement d'attitude vis-à-vis de la consommation de protéines animales. Les participants en contexte de compétition ont tendance à modifier leurs attitudes concernant la souffrance et la violence animal dans les abattoirs (M = 1.39 ; SD = 3.81 ; p < .05) : à la suite de leurs recherches d'information, ils tendent à moins considérer que les animaux ne devraient pas souffrir dans les abattoirs et que les éleveurs devraient s'engager contre la souffrance animale. À l'inverse, les participants en contexte de coopération ont tendance à renforcer leurs attitudes antérieures, c'est-à-dire à penser que les animaux ne devraient pas souffrir dans les abattoirs et que les éleveurs devraient s'engager contre la souffrance animal (M = -.86 ; SD = 1.82 ; p < .05). Précisons toutefois que l'écart-type du groupe « compétition » est particulièrement élevé par rapport au groupe « coopération », ce qui suggère des variations plus contrastées dans les changements d'attitudes au sein de ce groupe. Nos résultats montrent également que plus les participants croient en la coopérativité, plus ils renforcent leurs attitudes préexistantes pour devenir encore plus convaincus que les produits cosmétiques ne devraient pas être testés sur les animaux (E = -.250 ; t = -1.951 ; p = .058 ; $\chi^2 = 76.38$; p < .001). Les participants qui développent le plus une attitude positive à l'égard de la compétition renforcent quant à

[2] Précisons que si le modèle linéaire généralisé 1 du tableau 11 est significatif, cette significativité n'est pas pertinente pour notre étude dans la mesure où aucune variable du modèle ne contribue significativement à cette significativité générale et interdit donc toute forme de comparaison entre les effets de ces variables.

eux leurs attitudes vis-à-vis de la souffrance animale et de la violence dans les abattoirs : à la suite de leurs recherches d'information, ils sont encore plus convaincus par l'idée que les animaux peuvent souffrir dans les abattoirs et que les agriculteurs ne devraient pas être tenus de s'occuper de la souffrance animale (E = -.114, t = -2.626, p < .05, χ^2 = 68,81, p < .001). En ce qui concerne l'attitude à l'égard des questions politiques et écologiques, nos résultats montrent que la compétitivité et la coopérativité ont des effets similaires : plus les participants ont une attitude favorable à l'égard de la compétition et de la coopération, plus ils renforcent leurs convictions antérieures selon lesquelles l'agriculture intensive devrait être réglementée pour lutter contre le changement climatique[3] (pour la compétitivité ; E = -.172, t = -3,657, p < .01, χ^2 = 173,65, p < .001 ; pour la coopérativité ; E = -. 124, t = -2.328, p < .05, χ^2 = 173.65, p < .001).

Tableau 12. Mesures des effets de la compétition-coopération sur le changement-renforcement d'attitude[4]

	Compétition	**Coopération**	**Mesure**
Vd_9	1.65 (*SD* = 3.72)	1.24 (*SD* = 3.94)	$t_{(45)} = .371$; $p = .712$
Vd_10	1.85 (*SD* = 3.06)	.71 (*SD* = 3.58)	$t_{(39.5)} = 1.15$; $p = .257$
Vd_11	1.39 (*SD* = 3.81)	-.86 (*SD* = 1.82)	$t_{(37.5)} = 2.65$; $p = .012$
Vd_12	.58 (*SD* = 4.46)	.14 (*SD* = 2.85)	$t_{(43)} = -.405$; $p = .688$

[3] Le tableau 13 montre, d'une part, que le modèle linéaire généralisé 2 est significatif et, d'autre part, que chacune des variables considérées dans le modèle contribue significativement à cette significativité générale par le biais d'une corrélation négative.

[4] Notons que les écarts-types des variables 9 à 12 sont élevés en comparaison avec les moyennes.

Tableau 13. Mesures des effets de la compétitivité-coopérativité sur le changement-renforcement d'attitude

	E	*SE*	*t*	*p*	x^2	*p*
Vd_9. Modèle linéaire généralisé 1						
Coopérativité (cognition)	-.06180	.14239	-.434	.6666		
Coopérativité (comportement)	.05258	.15136	.347	.7301		
Coopérativité (affect)	.49403	.22073	2.238	.0308	131.92	<.001
Compétitivité (cognition)	-.22159	.14300	-1.550	.1291		
Compétitivité (comportement)	.18653	.20287	.919	.3634		
Compétitivité (affect)	-.09713	.13005	-.747	.4595		
Vd_9. Modèle linéaire généralisé 2						
Coopérativité	.10641	.05877	1.811	.077	67.63	<.001
Compétitivité	-.04499	.05177	-.869	.390		
Vd_1. Modèle linéaire généralisé 1						
Coopérativité (cognition)	-.25074	.12850	-1.951	.0580		
Coopérativité (comportement)	.24224	.13659	1.773	.0838		
Coopérativité (affect)	-.08148	.19919	-.409	.6847	76.38	<.001
Compétitivité (cognition)	-.14966	.12904	-1.160	.2530		
Compétitivité (comportement)	.03848	.18308	.210	.8346		
Compétitivité (affect)	-.01981	.11736	-.169	.8668		
Vd_1. Modèle linéaire généralisé 2						
Coopérativité	-.03249	.05351	-.607	.547	15.26	<.01
Compétitivité	-.05261	.04715	-1.116	.270		
Vd_11. Modèle linéaire généralisé 1						
Coopérativité (cognition)	-.14846	.12269	-1.210	.2334		
Coopérativité (comportement)	.13942	.13042	1.069	.2915		
Coopérativité (affect)	.07469	.19019	.393	.6966	96.77	<.001
Compétitivité (cognition)	-.20783	.12321	-1.687	.0994		
Compétitivité (comportement)	-.14060	.17480	-.804	.4259		
Compétitivité (affect)	-.01047	.11206	-.093	.9260		
Vd_11. Modèle linéaire généralisé 2						
Coopérativité	-.002195	.049479	-.044	.9648	68.81	<.001
Compétitivité	-.114470	.043591	-2.626	.0118		
Vd_12. Modèle linéaire généralisé 1						
Coopérativité (cognition)	-.259911	.130721	-1.988	.05365		
Coopérativité (comportement)	.044333	.138956	.319	.75136		
Coopérativité (affect)	-.099163	.202636	-.489	.62726	218	<.001
Compétitivité (cognition)	-.367985	.131274	-2.803	.00777		
Compétitivité (comportement)	.005154	.186245	.028	.97806		
Compétitivité (affect)	-.115509	.119394	-.967	.33913		
Vd_12. Modèle linéaire généralisé 2						
Coopérativité	-.12435	.05341	-2.328	.024566	173.65	<.001
Compétitivité	-.17209	.04706	-3.657	.000678		

5 Discussion

En situation de coopération, les participants ont tendance à s'engager dans des recherches d'informations qui sont axées sur l'exploitation navigationnelle de leurs requêtes et des résultats qui y sont associés. En comparaison aux participants en situation de compétition, les participants en situation de coopération sont moins susceptibles d'explorer et de réutiliser ces requêtes et ces résultats. Comparativement à la situation

de compétition, la situation de coopération encourage le développement de stratégies navigationnelles et thématiques qui sont plus pertinentes sur le plan épistémique. Contrairement aux participants en situation de compétition, les participants en situation de coopération montrent en effet une plus grande propension à développer :

- des connaissances factuelles sur le sujet de la consommation de protéines animales,
- une humilité épistémique qui se traduit par l'adoption d'une attitude plus positive, bienveillante moralement et donc acceptable socialement à l'égard des problèmes éthiques que pose ce même sujet (*cf.* la question de la souffrance animale).

Nos résultats montrent également que, par rapport à la compétitivité, la coopérativité conduit les participants à être moins sceptiques vis-à-vis de la bienveillance des sources qu'ils consultent, mais aussi à développer une humilité intellectuelle se traduisant par l'adoption d'attitudes plus vertueuses moralement et acceptables socialement concernant les problèmes éthiques susmentionnées (*cf.* la question de la souffrance animale).

6 Conclusion

Nos résultats montrent que par rapport à la compétition-compétitivité, la coopération-coopérativité conduit les participants à développer des buts de maîtrise qui favorisent le déploiement de stratégies d'exploitation navigationnelle qui sont plus pertinentes sur le plan épistémique. Ces stratégies permettent en effet à ces participants de développer une vigilance épistémique plus positive à l'égard des sources qu'ils consultent. Cette diminution du scepticisme peut évidemment entraîner des conséquences épistémiques négatives dans les situations de désinformation. Il n'en reste pas moins que la coopération-coopérativité encourage :

- le développement des connaissances factuelles des participants sur le sujet de la consommation de protéines animales ;
- le développement d'une humilité épistémique se traduisant par l'adoption d'attitudes de plus en plus tournées vers les positions dont les arguments sont les plus solides scientifiquement (*cf.* les questions éthiques liées à la souffrance animale).

References

1. Ames, C. (1992). Achievement goals and the classroom motivational climate. In D. H. Schunk & J. L. Meece (Eds.), *Student perceptions in the classroom* (pp. 327–348). Lawrence Erlbaum Associates, Inc.
2. Brante, E. W., & Strømsø, H. I. (2018). Sourcing in text comprehension: A review of interventions targeting sourcing skills. *Educational Psychology Review*, *30*(3), 773–799. https://doi.org/10.1007/s10648-017-9421-7
3. Bråten, I., Britt, M. A., Strømsø, H. I., & Rouet, J. (2011). The Role of Epistemic Beliefs in the Comprehension of Multiple Expository Texts: Toward an Integrated Model. *Educational Psychologist*, *46*(1), 48-70. https://doi.org/10.1080/00461520.2011.538647

4. Bråten, I., Stadtler, M., & Salmerón, L. (2017). *The role of sourcing in discourse comprehension.* In M. F. Schober, D. N. Rapp, & M. A. Britt (Eds.), Handbook of discourse processes (2nd. ed.). New York: Routledge.
5. Butera, F., & Mugny, G. (1995). Conflict between incompetences and influence of a low-expertise source in hypothesis testing. *European Journal of Social Psychology, 12*(4), 457–462. https://doi.org/10.1002/ejsp.2420250408
6. Collins-Thompson, K., Rieh, S. Y., Haynes, C. C., & Syed, R. (2016). Assessing learning outcomes in web search. In *Proceedings of the Conference on Human Information Interaction and Retrieval (CHIIR'16),* pp. 163-172. https://doi.org/10.1145/2854946.2854972
7. Deutsch, M. (1949). A Theory of Co-operation and Competition. *Human Relations*, *2*(2), 129-152. https://doi.org/10.1177/001872674900200204
8. Dieguez, S., Delouvée, S. (2021). *Le complotisme. Cognition, culture et société*. Bruxelles : Mardaga supérieur.
9. Dosso, C., Chevalier, A., Tamine, L., Paubel, P., & Salmerón, L. (2024). People also ask: How does this tool affect exploration-exploitation strategies with regard to prior domain knowledge and search context? An eye-tracking study. *Applied Ergonomics*, *121*, 104367. https://doi.org/10.1016/j.apergo.2024.104367
10. Dosso, C., Tamine, L., Paubel, P., & Chevalier, A. (2023). Navigational and thematic exploration–exploitation trade-offs during web search: effects of prior domain knowledge, search contexts and strategies on search outcome. *Behaviour And Information Technology*, 1-27. https://doi.org/10.1080/0144929x.2023.2242514
11. Dweck, C. S. (1986). Motivational processes affecting learning. *American Psychologist, 41*(10), 1040–1048. https://doi.org/10.1037/0003-066X.41.10.1040
12. Gadiraju, U., Yu, R., Dietze, S., & Holtz, P. (2018). Analyzing Knowledge Gain of Users in Informational Search Sessions on the Web. In *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval (CHIIR '18)*, pp. 2-11. https://doi.org/10.1145/3176349.3176381
13. Golman, R., & Loewenstein, G. (2018). Information gaps: A theory of preferences regarding the presence and absence of information. *Decision, 5*(3), 143–164. https://doi.org/10.1037/dec0000068
14. Hardy, J. H. III, Day, E. A., & Arthur, W., Jr. (2019). Exploration-exploitation tradeoffs and information-knowledge gaps in self-regulated learning: implications for learner controlled training and development. *Human Resource Management Review, 29*(2), 196–217. https://doi.org/10.1016/j.hrmr.2018.07.004
15. Johnson, D.W, & Johnson, R.T. (1985) Motivational Processes in Cooperative, Competitive and Individualistic Learning Situations. In: Ames, C. and Ames, R., Eds., *Research on Motivation in Education, Vol. 2*, Academic Press, New York, 249-286.
16. Johnson, D.W, & Johnson, R.T. (1989). *Cooperation and competition: theory and research.* Edina, MN: Interaction Book Company.
17. Kaplan, A. (2004). Achievement goals and intergroup relations. In P. R. Pintrich & M. L. Maehr (Eds.), Advances in research on motivation and achievement: Vol. 13. Motivating students, improving schools: The legacy of Carol Midgley (pp. 97–136). Amsterdam: Elsevier.
18. Levy, I., Kaplan, A., & Patrick, H. (2004). Early adolescents' achievement goals, social status, and attitudes towards cooperation with peers. *Social Psychology of Education, 43*(2), 1–33.
19. Maehr, M. L. (1984). Meaning and Motivation: Toward a Theory of Personal Investment. *Research on Motivation in Education, 1*, 115-144.

20. Mason, L., Boldrin, A., & Ariasi, N. (2010). Searching the Web to learn about a controversial topic: Are students epistemically active?. *Instructional Science, 38*, 607-633.
21. Mercier, H., & Sperber, D. (2017). *The enigma of reason.* Harvard University Press. https://doi.org/10.4159/9780674977860
22. Nicholls, J. G. (1984). Achievement motivation: Conceptions of ability, subjective experience, task choice, and performance. *Psychological Review, 91*(3), 328–346. https://doi.org/10.1037/0033-295X.91.3.328
23. Palani, S., Ding, Z., MacNeil, S., & Dow, S. P. (2021). The « active search » hypothesis. In *Proceedings of the Conference on Human Information Interaction and Retrieval (CHIIR'21)*, pp. 325-329. https://doi.org/10.1145/3406522.3446046
24. Pardi, G., von Hoyer, J., Holtz, P., & Kammerer, Y. (2020). The role of cognitive abilities and time spent on texts and videos in a multimodal searching as learning task. In *Proceedings of the Conference on Human Information Interaction and Retrieval (CHIIR'20)*, pp. 378-382. https://doi.org/10.1145/3343413.3378001
25. Petty, R. E., & Cacioppo, J. T. (1986). *The elaboration likelihood model of persuasion*. Springer New York.
26. Poortvliet, P. M., & Darnon, C. (2010). Toward a more social understanding of achievement goals: The interpersonal effects of mastery and performance goals. *Current Directions in Psychological Science, 19*(5), 324–328. https://doi.org/10.1177/0963721410383246
27. Rouet, J.-F., & Tricot, A. (1998). Chercher de l'information dans un hypertexte : vers un modèle des processus cognitifs. In A. Tricot & J.-F. Rouet (Eds.), *Les hypermédias, approches cognitives et ergonomiques* (pp. 57-74). Paris : Hermès.
28. Salmerón, L., & Llorens, A. (2018). Instruction of digital reading strategies based on eye-movements modeling examples. Journal of Educational Computing Research, 57(2), 343-359. https://doi.org/10.1177/0735633117751605
29. Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G., & Wilson, D. (2010). Epistemic vigilance. *Mind & Language, 25*(4), 359-393. https://doi.org/10.1111/j.1468-0017.2010.01394.x
30. Stromso, H. I., Bråten, I., Britt, M. A., & Ferguson, L. E. (2013). Spontaneous sourcing among students reading multiple documents. *Cognition and Instruction, 31*(2), 176–203. https://doi.org/10.1080/07370008.2013.769994
31. Toma, C, Gilles, I, Butera, F. (2011). Strategic use of preference confirmation in group decision making: the role of competition and dissent. *Br J Soc Psychol*., 52(1), 44-63. https://doi.org/10.1111/j.2044-8309.2011.02037.x
32. Toma, C., Vasiljevic, D., Oberlé, D., Augustinova, M., & Butera, F. (2012). Le traitement motivé de l'information dans la prise de décision en groupe: Le cas du paradigme des «profils cachés» [Motivated information processing in group decision making: The case of hidden profiles]. *L'Année Psychologique, 112*(4), 663–693. https://doi.org/10.4074/S000350331200406X
33. Tricot, A., & Comtat, P. (2012). La recherche d'informations comme dialogue. *Les Cahiers du numérique, 8*(1), 15-38.
34. Tricot, A., Sahut, G., & Lemarié, J. (2016). *Le document : communication et mémoire*. Louvain-la-Neuve : De Boeck Supérieur

Prédiction du risque d'échec dans un MOOC : une approche basée sur l'analyse des séries temporelles multivariées

Anass El Ayady[1,2], Maxime Devanne[2], Germain Forestier[2], and Nour El Mawas[1]

[1] CREM, Université de Lorraine, Metz, France
{anass.el-ayady, nour.el-mawas}@univ-lorraine.fr
[2] IRIMAS, Université de Haute-Alsace, Mulhouse, France
{maxime.devanne, germain.forestier}@uha.fr

Résumé Les MOOCs offrent un accès libre et gratuit à un large public, mais les taux de complétion restent faibles, souvent en raison du manque de personnalisation. Pour y remédier, il est essentiel de prédire les performances des apprenants afin de leur fournir une rétroaction adaptée. Les traces comportementales (clics, événements) peuvent être analysées sous forme de séries temporelles pour anticiper leurs résultats. Ce travail compare des méthodes de classification de séries temporelles multivariées pour identifier les apprenants à risque d'échec à différents moments du cours (après 5, 10 semaines, etc.). L'évaluation expérimentale, menée sur la base *Open University Learning Analytics Dataset* (OULAD), porte sur trois cours : deux en **STEM** et un en **SHS**. Les résultats préliminaires montrent que les approches évaluées sont prometteuses pour prédire l'échec dans les MOOCs. Les analyses suggèrent aussi que la précision dépend du volume d'interactions, soulignant l'importance de données comportementales riches et variées.

Mots clés: MOOCs · Séries temporelles · Apprentissage automatique · Prédictions précoces · Personnalisation du contenu

Abstract MOOCs offer free and open access to a wide audience, but completion rates remain low, often due to a lack of personalized content. To address this issue, it is essential to predict learner performance in order to provide tailored feedback. Behavioral traces—such as clicks and events—can be analyzed as time series to anticipate learners' outcomes. This work compares multivariate time series classification methods to identify at-risk learners at different stages of the course (after 5, 10 weeks, etc.). The experimental evaluation, conducted on the Open University Learning Analytics Dataset (OULAD), focuses on three courses : two in STEM and one in SHS. Preliminary results show that the evaluated approaches are promising for predicting learner failure in MOOCs. The analysis also suggests that prediction accuracy is influenced by the amount of recorded interactions, highlighting the importance of rich and diverse behavioral data.

Keywords : MOOCs, Time series, Machine learning, Early predictions, Content personalization.

1 Introduction

Avec l'essor des technologies numériques, les MOOCs se sont imposés comme un outil clé de démocratisation de l'éducation. Ils permettent à des millions d'apprenants, de tous horizons, d'accéder à des contenus éducatifs variés. Pourtant, leur potentiel est souvent freiné par des taux d'abandon élevés qui varient généralement entre 0,7% et 52,1%, avec une médiane de 12,6% [5]. Cela souligne un défi persistant dans l'adaptation des cours aux besoins spécifiques des apprenants.

En parallèle, l'évolution des outils d'intelligence artificielle et des méthodes d'apprentissage automatique offre de nouvelles opportunités pour résoudre ces problématiques. Les approches modernes basées sur l'analyse des données permettent non seulement d'automatiser l'identification des facteurs de risque, mais également d'adopter une approche proactive pour engager les apprenants. En mobilisant des algorithmes sophistiqués, il devient possible de proposer des contenus adaptés et des stratégies pédagogiques personnalisées, maximisant ainsi l'expérience éducative.

L'intégration de méthodes d'analyse des données provenant des interactions des apprenants avec la plateforme pourrait constituer une réponse prometteuse à ce défi. Ces données, parfois négligées, représentent une véritable source d'informations sur les comportements et les trajectoires des apprenants au fil du temps. En particulier, les clics de souris peuvent être exploités pour modéliser des séries temporelles décrivant leurs interactions avec différentes activités pédagogiques. Ces séries permettent de suivre les rythmes d'apprentissage, d'identifier les tendances et de prédire les risques d'échec avant qu'ils ne deviennent critiques.

Ce travail se distingue par l'utilisation des séries temporelles multivariées, rarement exploitées dans ce contexte, et l'introduction du modèle *Fully Convolutional Networks* (FCN) pour la prédiction du risque d'échec dans les MOOCs. Par ailleurs, il propose une analyse différenciée selon le type de cours (STEM vs SHS), un aspect encore peu exploré dans la littérature. L'objectif est d'identifier les étudiants à risque dès les premières semaines afin de favoriser des interventions pédagogiques ciblées.

2 Cadre théorique

Les recherches sur la prédiction des performances des apprenants dans les environnements d'apprentissage en ligne s'appuient de plus en plus sur des approches issues de l'intelligence artificielle, de l'apprentissage automatique et de l'apprentissage profond. Le succès récent des modèles d'apprentissage profond motive leur adoption croissante pour exploiter les données disponibles, notamment dans des bases comme OULAD [9], reconnue pour sa richesse et son accessibilité.

Les modèles récurrents, tels que ceux utilisant des LSTM, se sont montrés efficaces pour capturer les dépendances temporelles dans les données issues d'OULAD. Par exemple, le modèle DOPP [7] a démontré son utilité pour identifier les apprenants à risque dès les premières semaines, ce qui le rend pertinent pour une intervention précoce. Une approche hybride combinant des couches denses et des LSTM a également été proposée pour la prédiction multiclasses, illustrant la capacité de ces modèles à traiter des tâches complexes. De plus, les modèles GRU et LSTM étudiés dans [3] et [1] confirment leur pertinence pour analyser les séquences temporelles et prédire les risques d'échec des apprenants, en particulier dans des contextes nécessitant des prédictions précoces et fiables.

D'autres méthodologies exploitent des relations complexes entre apprenants et cours. Une approche basée sur des graphes neuronaux, combinée à des modèles récurrents, représente ces relations sous forme de graphes de connaissances [6]. En utilisant des données de clics et des attributs relationnels temporels, cette méthode démontre son efficacité pour effectuer des prédictions précoces dans un contexte multi-relationnel. Ces travaux mettent en avant l'intérêt d'explorer des représentations graphiques pour améliorer la compréhension et la prédiction dans les systèmes éducatifs.

Les modèles basés sur des Transformers [8] se distinguent par leur capacité à capturer des dépendances séquentielles grâce à des mécanismes d'attention. Ces approches permettent de traiter des données normalisées et organisées de manière séquentielle pour des prédictions multiclasses. En exploitant des données agrégées à différentes granularités temporelles, les Transformers démontrent leur efficacité pour les tâches de classification séquentielle, surpassant souvent les modèles traditionnels comme les LSTM.

Des méthodes plus classiques exploitent les forêts de séries temporelles pour classer les apprenants en fonction de leurs comportements interactifs [2]. De manière similaire, des modèles d'apprentissage automatique plus simples, comme les forêts aléatoires, permettent des prédictions précoces en analysant une partie limitée des données disponibles. Ces travaux illustrent la robustesse des algorithmes classiques lorsqu'ils sont appliqués à des données bien structurées.

Comparativement, les modèles récurrents (LSTM, GRU) excellent dans l'analyse des données temporelles continues et les prédictions à différents moments d'un cours grâce à leur capacité à modéliser les séquences. Les approches par graphes et Transformers s'avèrent efficaces pour capturer des relations complexes entre apprenants et contenus. Les modèles classiques, tels que les forêts aléatoires, bien que robustes et simples, sont moins adaptés aux prédictions séquentielles ou aux contextes relationnels complexes.

Cependant, ces avancées ne prennent pas en compte l'influence des différences disciplinaires, notamment entre les cours en SHS et en STEM. Ces distinctions, pourtant essentielles, peuvent affecter la densité des données et les performances des modèles prédictifs. Notre étude comble cette lacune en analysant ces deux contextes distincts et en évaluant les approches pour chaque discipline.

Dans notre étude, nous avons sélectionné les modèles LSTM pour leur performance reconnue en analyse de séries temporelles, comme indiqué dans la littéra-

ture. Le modèle DOPP, basé sur LSTM, est réputé pour capturer efficacement les dépendances temporelles et fournir des prédictions précoces à partir des données de clics.

En complément, nous avons exploré le FCN, qui se distingue par sa capacité à extraire des motifs locaux et globaux dans les séries temporelles grâce à ses couches convolutives. Bien qu'ils soient moins présents dans les travaux de l'état de l'art sur les MOOCs, ils ont démontré leur efficacité dans la classification des séries temporelles dans d'autres domaines [4], ce qui justifie leur inclusion dans notre étude.

Pour compléter cette sélection, nous avons intégré des modèles classiques comme les k-plus-proches-voisins (KNN) et les perceptrons multicouches (MLP). Bien qu'ils ne soient pas couramment étudiés dans le domaine de la prédiction des performances dans les MOOCs, ces modèles offrent une base de comparaison pour évaluer la pertinence et la valeur ajoutée des approches avancées. L'objectif de cette étude comparative est d'évaluer différentes méthodes d'analyse des séries temporelles, dans des conditions cohérentes, afin d'identifier celles qui semblent les plus prometteuses pour prédire le risque d'échec dans les contextes ciblés. En particulier, les modèles avancés comme le FCN et le DOPP exploitent les dépendances temporelles et les séries multivariées, tandis que les KNN et les MLP permettent d'évaluer la performance de solutions plus simples. Ce choix équilibré entre approches classiques et avancées permet d'obtenir une vue d'ensemble complète et rigoureuse des méthodes disponibles pour ce contexte spécifique.

3 Méthodologie

Cette section détaille les approches méthodologiques adoptées pour analyser les séries temporelles des interactions des apprenants et prédire leur niveau de réussite. Nous décrivons d'abord le jeu de données utilisé, puis les étapes de préparation des données, les modèles évalués, ainsi que les méthodologies d'évaluation utilisées pour comparer leurs performances.

3.1 Description du jeu de données

Cette étude s'appuie sur le jeu de données public OULAD, qui regroupe les parcours de 32 593 étudiants répartis sur 22 cours en ligne, avec plus de 10 millions de clics enregistrés. Nous avons sélectionné trois cours fortement représentés (tableau 1) : (BBB, SHS), (DDD, STEM) et (FFF, STEM), afin d'explorer l'influence du type disciplinaire sur les performances prédictives.

Chaque cours est proposé deux fois par an (sessions de février "B" et d'octobre "J") (tableau 2). Les sessions 2013B et 2013J sont utilisées pour l'entraînement, et les sessions 2014B et 2014J pour le test, ce qui garantit une évaluation sur des cohortes indépendantes dans le temps. Certains apprenants inscrits ne participent pas au cours, ce qui explique la différence entre le nombre total d'apprenants indiqué dans le tableau 1 et la somme des apprenants par période dans le tableau 2.

Tableau 1. Statistiques des apprenants et des présentations pour chaque module

Code de module	Nombre de présentations	Nombre d'apprenants
AAA	2	748
BBB	4	7909
CCC	2	4434
DDD	4	6272
EEE	3	2934
FFF	4	7762
GGG	3	2534
Total	**22**	**32593**

Tableau 2. Description des cours étudiés

Code	Type	Période	Nombre d'apprenants
BBB	SHS	2013B	1537
BBB	SHS	2013J	1870
BBB	SHS	2014B	1294
BBB	SHS	2014J	1921
DDD	STEM	2013B	1214
DDD	STEM	2013J	1768
DDD	STEM	2014B	1116
DDD	STEM	2014J	1647
FFF	STEM	2013B	1510
FFF	STEM	2013J	2098
FFF	STEM	2014B	1363
FFF	STEM	2014J	2121

La distribution des résultats finaux est la suivante : *Pass* (42,55%), *Withdrawn* (24,62%), *Fail* (23,94%), *Distinction* (8,89%). Pour la classification binaire, *Pass* et *Distinction* sont regroupés en classe positive, tandis que *Fail* et *Withdrawn* forment la classe négative. Une baseline simple basée sur la classe majoritaire atteint un F1-score de 0,58 (binaire) et 0,48 (multiclasse), ce qui reste significativement inférieur aux performances obtenues par les modèles proposés.

3.2 Préparation des données

Les interactions des étudiants avec les cours sont transformées en séries temporelles multivariées, où chaque étudiant est décrit par une séquence hebdomadaire de clics répartis par type d'activité (forum, quiz, ressource, etc.). Chaque activité constitue un *canal*, soit une dimension distincte de la série.

La structure obtenue est $(n_samples, n_weeks, n_activities)$, où $n_samples$ est le nombre d'apprenants, n_weeks la durée d'observation (variant de 5 à 40 semaines) et $n_activities$ le nombre d'activités distinctes. Pour garantir l'homo-

généité des canaux, une normalisation Min-Max est appliquée individuellement sur chaque canal.

Les étiquettes de performance sont assignées selon deux cadres : *binaire* (Pass/Fail) et *multiclasse* (Distinction, Pass, Fail, Withdrawn), permettant d'évaluer les modèles à différents niveaux de granularité.

La Fig. 1 montre que les apprenants qui réussissent maintiennent une activité régulière tout au long du cours, tandis que la Fig. 2 illustre une baisse significative des interactions pour les apprenants en échec. Ces dynamiques confirment l'importance des traces d'interactions comme indicateurs précoces.

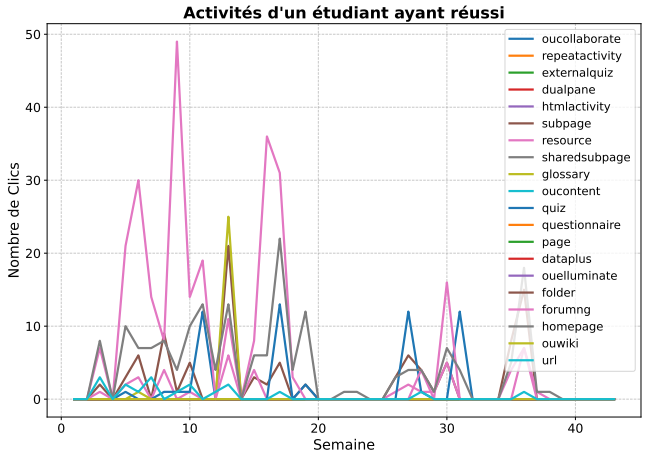


Fig. 1. Représentation des interactions hebdomadaires d'un apprenant ayant réussi. L'activité reste régulière tout au long du cours, indiquant un engagement constant.

3.3 Indicateur de performance

L'évaluation repose sur le *F1-score*, une métrique équilibrée combinant précision et rappel. Il est particulièrement adapté aux contextes de classification déséquilibrée, comme les MOOCs, où la simple précision globale (accuracy) peut être trompeuse.

Le F1-score est défini par :

$$F1 = \frac{2 \times \text{Précision} \times \text{Rappel}}{\text{Précision} + \text{Rappel}} \tag{1}$$

avec :

$$\text{Précision} = \frac{TP}{TP + FP}, \quad \text{Rappel} = \frac{TP}{TP + FN} \tag{2}$$

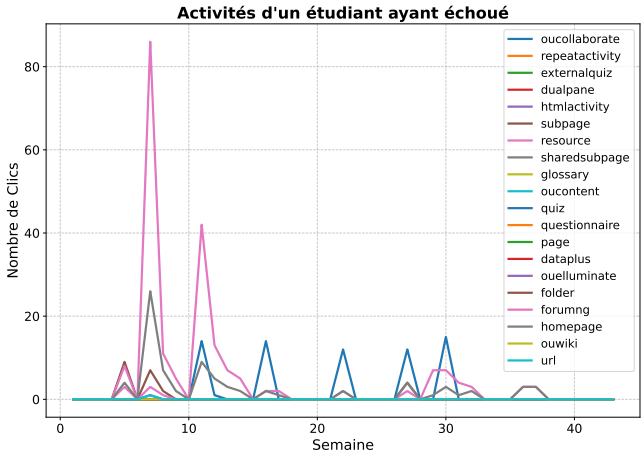


Fig. 2. Représentation des interactions hebdomadaires d'un apprenant échoué. Une baisse significative de l'activité est observée à partir de la 5e semaine, signalant un désengagement progressif.

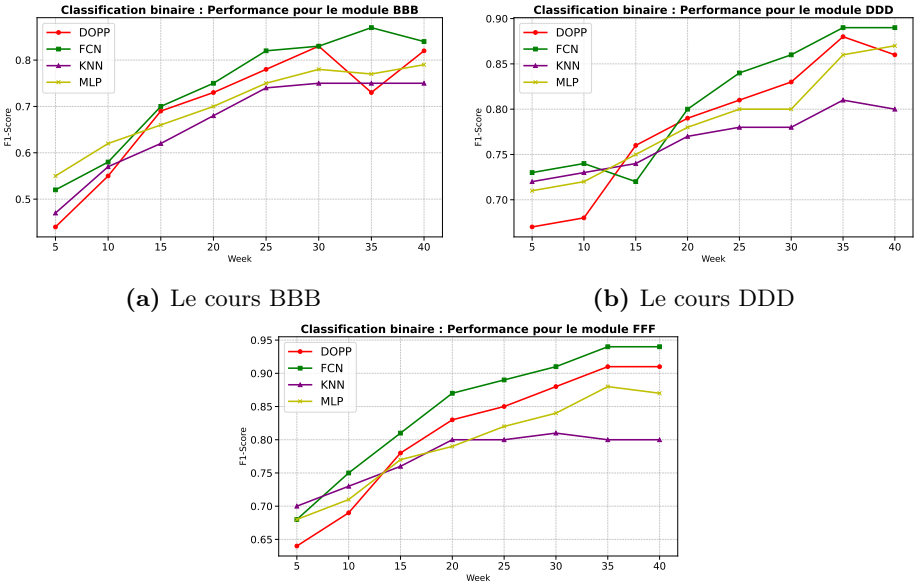
où TP, FP et FN désignent respectivement les vrais positifs, les faux positifs et les faux négatifs. Dans notre contexte, cela reflète la capacité du modèle à détecter correctement les apprenants à risque, tout en limitant les fausses alertes.

Les performances sont mesurées à différents moments du cours (après 5, 10, 15 semaines, etc.), afin d'évaluer la capacité des modèles à prédire tôt dans le parcours. Cette approche permet d'identifier les périodes critiques où les prédictions sont les plus fiables.

4 Résultats et Discussion

Les résultats obtenus pour la classification binaire et multiclasses révèlent des tendances similaires en termes de performances des modèles et de dépendance au type de cours et à la durée d'observation. Pour la **classification binaire** (Fig. 3), les cours **STEM** (**DDD** et **FFF**) affichent des scores F1 significativement plus élevés que le cours **SHS** (**BBB**), atteignant jusqu'à **94%** pour **FCN** après 40 semaines dans le cours FFF. Le cours **DDD**, bien qu'étant également un cours STEM, affiche des performances légèrement inférieures (86%). Cette différence s'explique par un volume et une régularité d'interactions inférieurs à ceux observés dans FFF, comme le confirment les statistiques d'usage. Ce résultat souligne la capacité du **FCN** à identifier des motifs temporels complexes grâce à ses convolutions, qui permettent d'extraire des relations locales et globales dans les séries temporelles tout en préservant leur structure séquentielle.

Cette supériorité des modèles pour les cours STEM s'explique sûrement par le volume élevé d'interactions et une variabilité temporelle plus marquée. Cela offre



(a) Le cours BBB **(b)** Le cours DDD

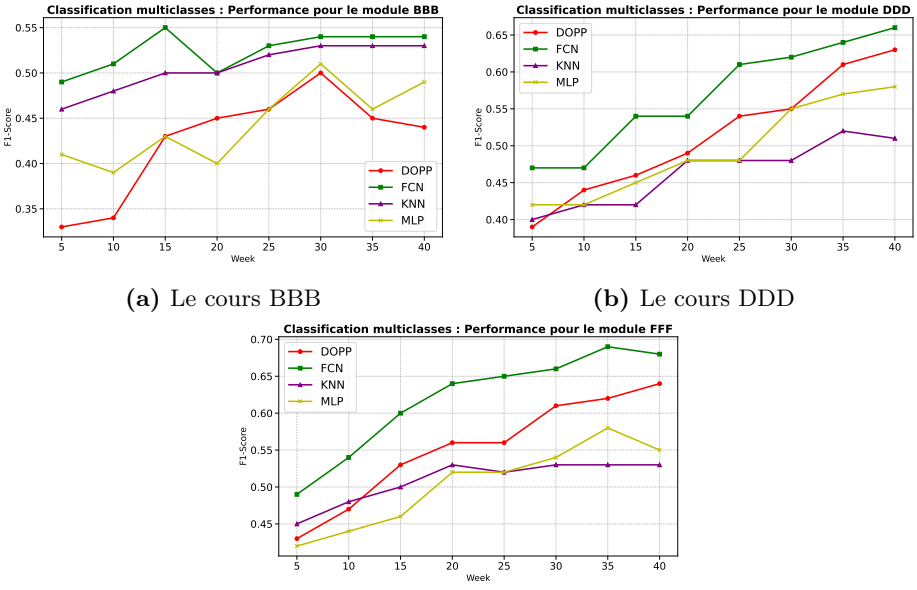
(c) Le cours FFF

Fig. 3. Évaluation de performances : classification binaire

aux modèles, notamment **FCN** et **DOPP**, une meilleure capacité d'apprentissage des relations complexes entre les activités et les résultats des apprenants. À l'inverse, le cours SHS souffre d'une stagnation des performances, en particulier pour les séries temporelles courtes, à cause d'une faible densité d'interactions : en configuration binaire, le F1-score reste limité à **58%** après 5 semaines ; en multiclasses, il débute à **32%** après 5 semaines et plafonne à **54%** après 40 semaines.

Pour la **classification multiclasses** (Fig. 4), ces tendances persistent, mais les scores F1 sont globalement plus faibles, en particulier pour le cours **BBB**, où **FCN** plafonne autour de **55%** après 40 semaines. Cela illustre la difficulté des modèles à discriminer plusieurs classes dans un environnement où les données interactives sont limitées. En revanche, pour les cours **DDD** et **FFF**, les modèles montrent une progression continue, avec **FCN** atteignant **66%** et **68%** respectivement. Cette amélioration s'accentue avec l'augmentation de la période d'observation, confirmant que l'enrichissement des données temporelles est crucial pour des prédictions précises, mais elle souligne aussi le défi du *cold start* pour les semaines initiales. Les modèles classiques comme le **KNN** stagnent dans les deux configurations de classification, prouvant leur faible capacité à capturer des dynamiques temporelles complexes.

Ainsi, que ce soit pour la classification binaire ou multiclasses, les cours présentant une richesse et une diversité élevées d'interactions temporelles offrent un avantage notable pour la prédiction des performances. À l'inverse, les cours



(a) Le cours BBB **(b)** Le cours DDD

(c) Le cours FFF

Fig. 4. Évaluation de performances : classification multiclasse

avec une faible densité d'interactions posent des défis importants. La performance supérieure des modèles neuronaux, comme **FCN** et **DOPP**, illustre leur capacité à apprendre efficacement des séries temporelles multivariées, même si des méthodes adaptées restent nécessaires pour améliorer les performances dans des contextes où les données sont limitées ou faiblement structurées, comme l'utilisation de modèles préentraînés ou d'apprentissage par transfert.

Dans les cours **SHS**, des stratégies complémentaires telles que l'intégration de données qualitatives, comme les réponses des apprenants dans les forums de discussion ou les commentaires sur les évaluations, ou d'autres indicateurs comportementaux, tels que le temps passé sur les activités, la participation aux sessions interactives ou l'utilisation des ressources pédagogiques, pourraient améliorer les performances des modèles.

La différence de performance observée entre les cours SHS et STEM peut être attribuée, en partie, à la conception pédagogique des cours. Une expérimentation complémentaire, illustrée dans (Fig. 5), montre que le cours **FFF** (STEM) présente un nombre unique d'activités et un total de clics nettement supérieurs à ceux du cours **BBB** (SHS). Cette différence suggère que les cours STEM bénéficient d'une conception favorisant une plus grande richesse et diversité d'interactions. Toutefois, cette observation ne peut être généralisée sans des analyses plus approfondies pour mieux comprendre l'impact des différences structurelles dans la conception pédagogique sur les performances des apprenants.

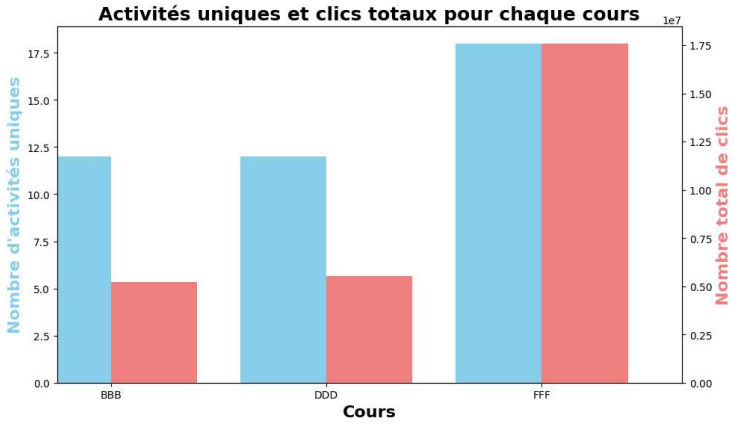


Fig. 5. Comparaison du nombre unique d'activités et du total de clics pour les cours SHS et STEM. Cette figure met en évidence que le cours STEM dispose de davantage d'activités et d'interactions, suggérant une conception pédagogique plus riche et variée.

5 Conclusion

En conclusion, cette étude met en évidence l'intérêt de l'analyse des séries temporelles multivariées pour classer les apprenants selon leur performance attendue dans les MOOCs. Les résultats montrent que la richesse et la diversité des interactions hebdomadaires jouent un rôle clé dans la précision des prédictions. Les modèles profonds, en particulier le FCN introduit ici, permettent de capturer des motifs temporels complexes, dépassant les approches classiques.

Ces travaux s'inscrivent dans la continuité des recherches en EIAH intégrant l'apprentissage automatique, avec une volonté d'innovation dans la personnalisation des parcours. À terme, des outils comme des systèmes d'analyse automatisée ou des tableaux de bord pour enseignants pourraient faciliter l'identification précoce des apprenants en difficulté et soutenir des interventions pédagogiques ciblées.

Cependant, cette étude présente certaines limites. Les conclusions sont basées sur un seul jeu de données (OULAD) et un nombre restreint de cours. Bien que ce dataset soit riche et structuré, des validations sur d'autres plateformes et contextes éducatifs sont nécessaires pour confirmer la généralisabilité des résultats.

Le problème du *cold start*, marqué par une faible densité d'interactions durant les premières semaines, reste un défi. Nos essais avec des tailles de noyaux de FCN et la distance *Dynamique Time Warping* (DTW) n'ont pas permis d'améliorations significatives. L'intégration d'informations contextuelles (telles que les profils démographiques et la date d'inscription) pourrait atténuer cet effet en apportant des indices supplémentaires lorsque les données comportementales sont insuffisantes.

Enfin, l'intégration de données qualitatives issues de forums, d'évaluations ou de retours libres représente une piste prometteuse pour enrichir les profils d'apprenants. Leur usage soulève cependant des défis méthodologiques liés à leur nature interprétative et déclarative. Des approches hybrides combinant données quantitatives et qualitatives, renforcées par des techniques comme le transfert d'apprentissage [10], pourraient améliorer la robustesse des analyses et leur adaptation à une diversité de contextes éducatifs.

Remerciements

Cette étude s'inscrit dans le projet **COPCOT** (ANR-22-CE38-0003), soutenu par l'Agence Nationale de la Recherche (**ANR**), visant à personnaliser les contenus éducatifs numériques pour faciliter l'acquisition de compétences en informatique. Nous remercions l'**ANR** pour son soutien financier et les partenaires du projet pour leurs contributions.

Références

1. Aljohani, N.R., Fayoumi, A., Hassan, S.U. : Predicting at-risk students using clickstream data in the virtual learning environment. Sustainability **11**(24), 7238 (2019)
2. Haiyang, L., Wang, Z., Benachour, P., Tubman, P. : A time series classification method for behaviour-based dropout prediction. In : 2018 IEEE 18th international conference on advanced learning technologies (ICALT). pp. 191–195. IEEE (2018)
3. He, Y., Chen, R., Li, X., Hao, C., Liu, S., Zhang, G., Jiang, B. : Online at-risk student identification using rnn-gru joint neural networks. Information **11**(10), 474 (2020)
4. Ismail Fawaz, H., Forestier, G., Weber, J., Idoumghar, L., Muller, P.A. : Deep learning for time series classification : a review. Data mining and knowledge discovery **33**(4), 917–963 (2019)
5. Jordan, K. : Massive open online course completion rates revisited : Assessment, length and attrition. International Review of Research in Open and Distributed Learning **16**(3), 341–358 (2015)
6. Karimi, H., Derr, T., Huang, J., Tang, J. : Online academic course performance prediction using relational graph convolutional neural network. International Educational Data Mining Society (2020)
7. Karimi, H., Huang, J., Derr, T. : A deep model for predicting online course performance. Association for the Advancement of Artificial Intelligence pp. 1–6 (2020)
8. Kusumawardani, S.S., Alfarozi, S.A.I. : Transformer encoder model for sequential prediction of student performance based on their log activities. IEEE Access **11**, 18960–18971 (2023)
9. Kuzilek, J., Hlosta, M., Zdrahal, Z. : Open university learning analytics dataset. Scientific data **4**(1), 1–8 (2017)
10. Tsiakmaki, M., Kostopoulos, G., Kotsiantis, S., Ragos, O. : Transfer learning from deep neural networks for predicting student performance. Applied Sciences **10**(6), 2145 (2020)

Formation aux compétences douces : vers un outil auteur pour la création d'agents conversationnels basés sur des LLM

Yassine El Boudouri[0009−0001−4293−961X], Walter Nuninger[0000−0002−2639−1359], Frédéric Hoogstoel[0000−0001−9179−9046], Luigi Lancieri[0000−0002−1406−2937], Julian Alvarez[0000−0002−9862−9485], and Yvan Peter[0000−0002−1145−357X]

Univ. Lille, CNRS, Centrale Lille, UMR 9189 CRIStAL, F-59000 Lille, France
prenom.nom@univ-lille.fr

Résumé. Les jeux de rôles constituent un excellent moyen de développer les compétences douces dans un environnement contrôlé. L'intégration d'agents conversationnels basés sur l'IA générative permet de les mettre en œuvre avec une grande diversité. Cet article présente un outil auteur conçu pour encadrer la création de prompts, visant à définir les besoins et la personnalité de clients dans le cadre de formations à la relation client. Il expose également les résultats d'une première expérimentation menée auprès d'enseignant·es utilisant cet outil.

Mots-clés : compétences douces · relation client · grand modèle de langue · ingénierie de prompt · agent conversationnel

Abstract. Role-playing games are excellent ways to develop soft skills in a controlled environment. The integration of generative AI-based conversational agents allows for their implementation with great diversity. This article presents an authoring tool designed to guide the creation of prompts aimed at defining the needs and personality of clients within the context of customer relationship training. It also outlines the results of an initial experiment conducted with teachers using this tool.

Keywords : soft skills · customer relationship · large langage models · prompt engineering · chatbot

1 Introduction

Le projet RVRC4.0 est un projet financé via l'appel d'offre DEFFINUM de 2022. Il vise à explorer l'utilisation des technologies numériques pour la formation aux compétences douces, essentielles aux métiers liés à la relation client, tels que vendeurs, conseillers bancaires, etc. L'objectif principal est de développer des ressources pédagogiques mobilisant des environnements virtuels intégrables dans les parcours de formation. Le projet cherche à évaluer leurs apports à l'amélioration de l'expérience d'apprentissage, en particulier par rappport la motivation et aux compétences.

La communication, la résolution de problèmes ou la gestion du temps et l'intelligence sociale et collective sont différentes facettes des compétences douces (*soft skills*) [8]. Elles sont à la fois difficiles à acquérir et à mesurer [6]. Bien que largement étudiées dans des contextes comme la santé ou l'ingénierie, elles restent peu explorées dans le domaine de la relation commerciale [4]. Elles sont pourtant fondamentales dans le secteur des services en contribuant directement à la satisfaction et à la fidélisation des clients [2].

Les pédagogies actives basées sur des mises en situation, via des jeux de rôles, sont utilisées dans les métiers de contact pour l'acquisition des compétences douces, notamment dans le milieu médical [12]. L'intelligence artificielle générative, et plus particulièrement les grands modèles de langage (LLM), ont le potentiel de générer des dialogues variés permettant une mise en situation réaliste. Ainsi, l'apprenant, incarnant le rôle d'un « vendeur, conseiller, hôte de caisse, etc. », se retrouverait confronté à un « client » dont les réponses sont générées en fonction de ses paroles. Cette variabilité enrichit la situation et, par conséquent, le scénario. L'apport de l'enseignant résiderait alors dans la définition de la situation, visant un objectif pédagogique précis.

Pour que les LLM soient utilisés efficacement dans un contexte pédagogique, ils doivent être correctement « conditionnés ». Cela demande de leur donner des instructions sous la forme de prompts clairs qui permettent de satisfaire des règles définies, d'adopter des comportements qui soient conformes, et ainsi fournir des réponses appropriées. La création de tels prompts repose sur un processus appelé « ingénierie des prompts », qui est une approche itérative et expérimentale nécessitant une bonne compréhension du fonctionnement des LLM. Cette tâche peut s'avérer difficile pour les enseignant·es non expert·es en intelligence artificielle.

Dans cet article, nous présentons une première version d'un outil auteur, à destination des enseignant·es, facilitant la définition et le test de prompts, soit l'édition de la situation et du but d'un agent jouant le rôle d'un client dans le cadre d'un échange commercial. Nous abordons dans un premier temps les usages des agents conversationnels dans le cadre de jeux de rôles avant de présenter le prototype élaboré. Nous terminons par une analyse des forces et des limites de la solution proposée avant de dresser les perspectives.

2 Jeux de rôles et agents conversationnels

Les jeux de rôle offrent aux apprenant·es l'opportunité d'une expérience proche de la réalité dans laquelle ils ou elles peuvent s'exercer sans risques. Elles les impliquent dans un apprentissage actif dans lequel les émotions et la réflexion sont propres à ancrer l'apprentissage [3]. L'utilisation d'un agent conversationnel dans un jeu de rôle permet à l'apprenant de travailler ses compétences via l'interaction avec un client simulé. Sun et al. (2024) ont exploré l'intérêt, l'usage et les limites de la définition et de l'utilisation de *personna* dans le configuration des agents conversationnels basée sur l'IA générative [11]. Le persona – qui est une représentation de personnalité propre à un agent – incorpore une per-

sonnalité spécifique rendant l'échange plus réaliste et permettant de confronter l'utilisateur à des comportements différents. Les auteurs soulèvent toutefois les problèmes du maintien de la consistance – capacité à maintenir sa personnalité – du persona qui n'est pas assurée actuellement par les modèles de langage et de l'évaluation de cette consistance.

Pester et al. (2024) ont passé en revue différents usages combinant grands modèles de langages et réalité virtuelle dont l'apprentissage des langues et le domaine médical (e.g. interaction avec les patients) [9]. Les auteurs relèvent la sensibilité des modèles de langages à de subtiles variations du prompt qui implique un travail de mise au point pour un résultat correct.

Ces différents travaux illustrent l'intérêt d'intégrer des agents conversationnels dans des jeux de rôles pour la pratique des compétences douces. Les LLM permettent une génération des dialogues de l'agent de manière adaptée à l'évolution du contexte. Les LLM ont démontré leur capacité à fournir des réponses conformes aux attentes humaines [7] mais aussi à simuler des sous-populations humaines spécifiques [1], et à faire preuve de capacités de prise de décision [5].

Ces travaux montrent toutefois l'importance du conditionnement via le prompt pour obtenir un résultat intéressant et incitent à être attentifs aux maintient de la consistance du rôle attribué à l'agent. Des propositions telles que DINA (*DIalogue for Natural interaction with intelligent Assistants*) [10] proposent un cadre pour la structuration des prompts mais celui-ci reste à la charge de l'utilisateur. Nous cherchons à proposer un outil pour simplifier cette production dans le contexte qui nous intéresse.

3 Élaboration d'un outil auteur

Nous visons à concevoir un outil auteur qui simplifie la création et l'expérimentation d'agents virtuels intelligents par les enseignant·es. Au centre de ce dispositif se trouve le *prompt système* : une instruction prédéfinie, saisie par l'usager, et transmise au modèle de langage pour orienter la génération des réponses.

Dans un usage générique, ce prompt peut se limiter à une formule simple, telle que : « Vous êtes un assistant utile. » Toutefois, lorsqu'il s'agit de faire incarner un rôle spécifique à un agent conversationnel, le prompt système doit être enrichi. Il devient alors essentiel d'y inclure des indications précises sur le persona à adopter, le ton à employer, ainsi que sur le contexte d'intervention.

Afin de faciliter l'appropriation par des enseignant·es non spécialistes de l'intelligence artificielle, nous proposons un outil auteur qui vise à équilibrer simplicité d'utilisation et précision des instructions données au modèle. L'objectif final est de permettre la création de scénarios où les dialogues générés sont non seulement adaptés au contexte et aux objectifs d'apprentissage, mais aussi suffisamment originaux pour maintenir l'attention des apprenants.

3.1 Prototypage de notre proposition

Un groupe de 26 enseignant·es de la région des Hauts-de-France a été identifié et sollicité pour participer à des ateliers dans le cadre du projet. Ces enseignant·es, spécialisé·es dans la relation client, interviennent dans diverses filières éducatives, notamment les CAP, les BAC professionnels et les BTS. Leurs domaines d'expertise incluent la banque, l'assurance, la vente, ainsi que d'autres secteurs où l'interaction directe avec les clients est essentielle.

Suite aux interactions avec les enseignant·es sur leurs enjeux et leurs besoins, nous avons élaboré un premier prototype présenté figure 1. Celui-ci comprend une zone de saisie du prompt de conditionnement et une zone de test de dialogue. Cette dernière fournit un chat où les enseignant·es peuvent interagir avec le personnage qu'ils ont créé et évaluer la conformité du persona et du scénario. La partie prompt est volontairement peu élaborée de manière à permettre une libre expérimentation par les enseignant·es de l'effet du prompt sur le dialogue généré et afin de pouvoir collecter les différents essais pour les analyser.



Figure 1. Interface de conditionnement et d'interaction avec un personnage.

Pour le modèle de langage, nous utilisons GPT-4o d'OpenAI, version 2024-08-06. En interne, le nom et la description de chaque personnage ont été combinés en un prompt système unique, formaté comme suit :

```
Votre nom est : {nom}
Vous jouez le rôle de : {description du personnage et de son besoin}
Gardez vos réponses courtes.
```

Il est complété par l'instruction « Gardez vos réponses courtes », afin d'éviter les réponses explicatives trop longues, qui pourraient paraître artificielles ou inadaptées au contexte. À chaque interaction, le modèle reçoit ce prompt système, enrichi par l'historique complet de la conversation, pour générer une réponse cohérente, contextualisée et alignée avec le rôle du personnage.

3.2 Expérimentation

Au cours d'un atelier en juin 2024, les enseignant·es ont testé librement l'outil proposé. L'outil a permis, avec leur consentement, de collecter des scénarios pédagogiques et d'enregistrer les tests de conversation réalisés. Compte tenu de leur domaine d'expertise et des besoins pédagogiques associés, tous les personnages créés pour les tests étaient modélisés comme des clients interagissant avec un apprenant-vendeur.

À l'issue des ateliers, les enseignant·es ont complété un questionnaire visant à évaluer leur perception de l'outil, tout en identifiant ses atouts et ses limites. Une question ouverte clé leur a été posée : « Quelles options pourraient être envisagées pour améliorer la création d'un personnage ? »

Les objectifs de cette première expérimentation étaient de valider l'intérêt pédagogique du système de dialogue avec l'agent conversationnel, de vérifier l'utilisabilité de l'interface et de mieux comprendre l'usage potentiel par les enseignant·es à travers leurs tests.

3.3 Résultats

Notre première analyse s'est concentrée sur l'examen des situations, des rôles créés et l'identification des éléments répétitifs dans les descriptions des personnages. Voici des exemples de personnages créés par les enseignant·es :

Brigitte
Dame de 65 ans , retraitée , qui souhaite acquérir son premier smartphone. Elle va voir un vendeur chez Auchan, elle décrit sa situation. Elle a enfants et petits-enfants, elle est engagée dans le milieu associatif. Le vendeur doit lui identifier ses besoins. Son budget est restreint, elle a peur de la technologie et veut faire plaisir à sa famille qui la pousse à acheter pour échanger avec elle sur différents réseaux car ils habitent loin.

Thierry
Homme de 48 ans passionné de métal avec trois enfants. Il vient chercher des places pour le hellfest. On lui a dit au téléphone qu' il restait quelques places. Le client vient de loin et se déplace exprès. Il ne veut pas être déçu d' être vnu pour rien. Il a un perroquet sur l' épaule et un bandeau sur l'oeil. Il ne voit pas bien et entend peu.

Madame Martin
Madame Marteens est une cliente fidele, discrete et reservee, de notre magasin en pret a porter. Cette cliente est d origine anglaise, a 40 ans , fraichement divorcee, maman de 3 enfants, de profession cadre en recherche d'emploi et reside a 30 kms du point de vente situe a Lille. La cliente recherche une tenue pour un entretien d'embauche.

Nous avons identifié quatre éléments récurrents que nous mettons en évidence de manière colorée dans les exemples ci-dessus : le genre en bleu, l'âge en violet, l'histoire du personnage en orange et les objectifs du personnage en vert. Ces éléments étaient systématiquement présents, en tout ou partie, dans la description des personnages créés (85 % les contenaient tous). Tous les personnages créés contiennent au moins deux éléments. Nous les considérons donc comme essentiels pour incarner un client « crédible » pour la situation pédagogique à créer parce que permettant un certain degré de réalisme. De plus, ces éléments suggèrent que, lorsqu'un jeu de rôle est utilisé pour enseigner les compétences comportementales dans le cadre du service client, les enseignant·es connaissent les panels de profils clients pouvant être rencontrés dans le cadre du métier. Par conséquent, ces éléments sont au cœur des objectifs de formation spécifiques.

Les enseignant·es avaient la liberté de décrire les personnages sans directives précises, ce qui a pu conduire à l'omission de certains aspects jugés moins essentiels ou non explicitement suggérés. En plus des éléments principaux identifiés, d'autres options, moins souvent mentionnées mais relevées par certain·es enseignant·es dans le questionnaire, incluent le niveau de langue, l'état d'esprit, l'état émotionnel et les ressources spécifiques du personnage (par exemple, un ticket de caisse ou une carte de fidélité). Intégrer certains ou tous ces éléments supplémentaires dans le processus de création pourrait significativement enrichir les résultats, en aidant les enseignant·es à se rappeler des aspects qu'ils auraient autrement négligés ou jugés secondaires. Notre méthode plus structurée pour définir les caractéristiques initiales des personnages permet ainsi de concevoir

un prompt qui favorise un développement plus riche et complet, tout en évitant de complexifier inutilement l'outil auteur.

Pour la prochaine itération d'amélioration de l'outil, nous avons donc décidé d'incorporer les éléments supplémentaires indiqués dans le Tableau 1.

Élément	**Valeurs Possibles**	**But**
Nom	Texte libre	Le nom du personnage.
Genre	Sélection dans une liste (ex. : Homme, Femme, Non-binaire)	Assurer l'inclusivité et la pertinence dans des situations de jeu en permettant des rôles réalistes.
Âge	Nombre (6 à 120)	Assurer le représentativité générationnelle par l'âge du personnage en années.
Histoire	Texte libre	Fournit un passé à l'agent qui traduit les spécificités personnelles ou professionnelles et ainsi, ajoute de la profondeur et clarifie les modes d'interaction.
Objectifs	Texte libre	Poser les motivations d'interaction de l'agent, et orienter l'exercice de formation pour l'attendu pédagogique.
Niveau de langue	Sélection dans une liste (ex. : Courant, Intermédiaire, Débutant)	Reflète le niveau de confort du personnage avec la langue, introduisant d'éventuels défis de communication dans le jeu de rôle.
Ressources	Texte libre (ex. : Ticket, Carte de fidélité, Reçu manquant)	Spécifie les objets ou outils possédé par le personnage, enrichissant le scénario et créant des opportunités de résolution de problèmes.

Table 1. Éléments pour créer un personnage

4 Discussion

L'ingénierie des prompts est un processus itératif nécessitant de nombreuses tentatives et expérimentations pour atteindre le résultat souhaité. Prenons l'exemple du personnage de « Madame Martin », décrit précédemment : l'enseignant qui a conçu ce personnage a remarqué que celui-ci, après avoir établi un budget initial, mais l'avait ensuite dépassé (voir le dialogue de test en Annexe A). En réponse, l'enseignant a ajusté la description du personnage, modifiant ainsi le prompt pour corriger ce comportement. Ce processus illustre le rôle de l'enseignant en tant qu'ingénieur de prompts.

L'expérimentation a montré que les enseignant·es ont rapidement adopté cette démarche, avec 64 % des scénarios créés qui ont été modifiés en cours de route pour affiner les comportements du modèle. Cependant, bien que les modèles de langage aient démontré leur potentiel pour incarner efficacement des

rôles assignés, certaines incohérences persistent, déjà relevées dans l'état de l'art (hallucinations, biais, etc.). Ces incohérences incluent des oublis dans le respect du rôle, des erreurs dans la prise de décision, une expression inadéquate des émotions requises, ainsi que des réponses répétitives, qui diminuent la qualité des interactions.

Un point particulièrement intéressant relevé par les enseignant·es concerne le comportement du modèle face à certaines entrées. Par exemple, lorsqu'une interaction commence par un simple « Bonjour », le modèle tend à répondre de manière standardisée par : « Bonjour ! Comment puis-je vous aider aujourd'hui ? » (prenant ainsi le rôle de vendeur). De plus, si une question hors contexte, telle que « Donnez-moi une définition du métabolisme », lui est posée, le modèle répond comme un assistant général plutôt que de rester dans son rôle assigné. Ces observations ont conduit à des ajustements dans les prompts pour minimiser ces problèmes.

Voici le prompt final résultant des analyses qualitatives et quantitatives :

```
Vous appelez {nom}.
Vous êtes {genre}.
Vous avez {âge} ans.
Vous êtes {histoire}.
Vous voulez {objectif}.
Vous avez à votre possession: {ressource}.

Consignes :
- Répondez de manière concise.
- Ne répondez pas par "Bonjour ! Comment puis-je
  vous aider aujourd'hui ?" après un "bonjour".
- Ne sortez jamais de votre personnage.
- Utilisez un niveau de langue {niveau_de_langue}.
```

Ce prompt est plus structuré et inclut des consignes précises pour guider le modèle. Il introduit une flexibilité en permettant d'inclure uniquement les informations fournies (seuls le nom et l'objectif sont obligatoires). Si un élément n'est pas précisé, il n'est pas inclus dans le prompt final. Ces consignes visent à réduire les incohérences dans les interactions et à garantir des réponses naturelles et pertinentes tout en maintenant un strict respect du rôle assigné.

Malgré les erreurs initiales, les enseignant·es ont exprimé des retours très positifs sur le potentiel de cet outil. Tous estiment qu'il est facile à utiliser, et à l'issue de l'atelier, 87 % d'entre eux envisagent de l'intégrer dans leurs cours. Les évolutions apportées à l'interface dédiée à l'élaboration des prompts visent à enrichir la définition des situations et des rôles, tout en réduisant les limitations observées.

5 Conclusion et perspectives

L'objectif du projet RVRC 4.0 est de développer un environnement virtuel configurable (boutique, agence...) permettant de mettre en situation les appre-

nants via une interaction avec un avatar piloté par un agent conversationnel basé sur un modèle de langage. Pour l'objectif d'apprentissage, le persona (profil à incarner par l'agent) et le contexte d'intervention doivent être prescrits. Cette responsabilité relève des enseignant·es en fonction de leurs objectifs pédagogiques. À ce stade, notre contribution principale réside dans la conception d'un outil auteur facilitant la création de prompts répondant à leur attentes. Ce processus, relevant de l'ingénierie des prompts, fait l'objet d'une amélioration continue grâce à la collaboration avec un groupe d'une trentaine d'enseignant·es pour faciliter leur appropriation et l'intégration dans leurs formations aux métiers de la vente et du conseil, couvrant les niveaux CAP, Bac professionnel et BTS.

Ce premier travail a permis de faire évoluer l'interface et le modèle de prompt du prototype. Une maquette de la nouvelle interface a été diffusée auprès des enseignant·es avec un questionnaire pour recueillir leurs réactions et questions. Le prototype va être rendu accessible aux enseignant·es pour qu'ils et elles puissent créer des clients à partir de prompts et tester des dialogues avec ceux-ci. L'instrumentation du prototype nous permettra d'avoir une trace des différents prompts définis et dialogues générés ce qui nous permettra une analyse plus exhaustive des éléments utilisés dans les prompts (histoire du personnage, but...) et des changements qu'ils opèrent en fonction du déroulement des dialogues générés. En parallèle, un questionnaire lié à chaque essai nous donnera un retour sur le fait que le dialogue généré est adapté à leurs attentes pédagogiques.

Références

1. Lisa P Argyle, Ethan C Busby, Nancy Fulda, Joshua R Gubler, Christopher Rytting, and David Wingate. Out of one, many : Using language models to simulate human samples. *Political Analysis*, 31(3) :337–351, 2023.
2. Barbara Cimatti. Definition, development, assessment of soft skills and their role for the quality of organizations and enterprises. *International Journal for quality research*, 10(1) :97, 2016.
3. Timothy C Clapper. Role Play and Simulation Returning to Teaching for Understanding. *Clapper, T. C. (2010). Role play and simulation. The Education Digest*, 75(8) :39–43, 2010.
4. Lorena C Espina-Romero, Sandra Lucia Aguirre Franco, Helga Ofelia Dworaczek Conde, Jesús M Guerrero-Alcedo, Doile Enrique Ríos Parra, and Juan Carlos Rave Ramírez. Soft skills in personnel training : Report of publications in scopus, topics explored and future research agenda. *Heliyon*, 9(4), 2023.
5. Shibo Hao, Yi Gu, Haodi Ma, Joshua Jiahua Hong, Zhen Wang, Daisy Zhe Wang, and Zhiting Hu. Reasoning with language model is planning with world model. *arXiv preprint arXiv :2305.14992*, 2023.
6. James J Heckman and Tim Kautz. Hard evidence on soft skills. *Labour economics*, 19(4) :451–464, 2012.
7. John J Horton. Large language models as simulated economic agents : What can we learn from homo silicus ? Technical report, National Bureau of Economic Research, 2023.

8. Miriam L Matteson, Lorien Anderson, and Cynthia Boyden. "soft skills" : A phrase in search of meaning. *portal : Libraries and the Academy*, 16(1) :71–88, 2016.
9. Andreas Pester, Ahmed Tammaa, Christian Gütl, Alexander Steinmaurer, and Samir Abou El-Seoud. Conversational Agents, Virtual Worlds, and Beyond : A Review of Large Language Models Enabling Immersive Learning. In *2024 IEEE Global Engineering Education Conference (EDUCON)*, pages 1–6, Kos Island, Greece, May 2024. IEEE.
10. Antônio Victor Figueiredo Porto and Maria Elizabeth Sucupira Furtado. Framework to Specify Dialogues for Natural Interaction with Conversational Assistants Applied in Prompt Engineering. In Kohei Arai, editor, *Intelligent Systems and Applications*, volume 1068, pages 231–253. Springer Nature Switzerland, Cham, 2024.
11. Guangzhi Sun, Xiao Zhan, and Jose Such. Building Better AI Agents : A Provocation on the Utilisation of Persona in LLM-based Conversational Agents. In *Proceedings of the 6th ACM Conference on Conversational User Interfaces*. arXiv, 2024.
12. Azzouzi Widad and Gantare Abdellah. Strategies used to teach soft skills in undergraduate nursing education : a scoping review. *Journal of Professional Nursing*, 42 :209–218, 2022.

A Exemple de dialogue

Ce rôle a été conçu par un enseignant spécialisé en relation client et est joué par GPT-4o. L'enseignant, quant à lui, joue le rôle du vendeur.

Rôle joué par le modèle de langue : Madame Marteens est une cliente fidèle, discrète et reéervée, de notre magasin en prêt a porter. Cette cliente est d'origine anglaise, a 40 ans, fraîchement divorcée, maman de 3 enfants, de profession cadre en recherche d'emploi et réside a 30 kms du point de vente situé à [anonyme] La cliente recherche une tenue pour un entretien d'embauche.

Vendeur : bonjour, qu'est-ce que tu veux
Cliente : Bonjour, je suis Madame Marteens. Je cherche une tenue pour un entretien d'embauche, s'il vous plaît !
Vendeur : bienvenue madame marteens
Cliente : Merci beaucoup ! Vous pourriez m'aider à trouver quelque chose qui soit à la fois professionnel et confortable ?
Vendeur : avez vous une idée de ce que vous chercher
Cliente : J'aimerais quelque chose d'élégant mais pas trop strict, peut-être un tailleur ou une robe avec une veste. Vous avez des suggestions ?
Vendeur : Quelles sont vos couleurs
Cliente : Je préfère les couleurs sobres, comme le bleu marine, le gris ou le noir.
Vendeur : quel est votre taille
Cliente : Je fais du 38, s'il vous plaît.
Vendeur : avez vous une idée de budget
Cliente : Oui, j'aimerais ne pas dépasser les 150 euros.
Vendeur : dans ce budget je peux vous proposer un pantalon

Cliente : D'accord, est-ce que vous avez un modèle élégant qui irait bien pour un entretien ?
Vendeur : je vous propose ce modèle a 140 euros, qu en pensez vous
Cliente : Ça semble bien, pourriez-vous me dire si c'est confortable et adapté pour un entretien ?
Vendeur : je vous propose de l'essayer avec ce chemisier
Cliente : D'accord, je vais essayer. Le chemisier est dans quelles couleurs ?
Vendeur : beige
Cliente : Parfait, je vais essayer l'ensemble, merci !
Vendeur : cet ensemble vous convient parfaitement
Cliente : Super, merci beaucoup ! Je vais le prendre alors.
Vendeur : je vous propose ce petit carre hermes pour compléter la tenue
Cliente : Bonne idée, ça apporterait une touche élégante ! Combien coûte-t-il ?
Vendeur : il est en promotion a 299 euros
Cliente : C'est un peu hors budget pour moi, je vais rester sur l'ensemble, merci quand même !
Vendeur : très bien, je vous prie de me suivre en caisse
Cliente : Merci, je vous suis !
Vendeur : le pantalon à 140 euros et la chemise à 85 euros, cela vous fera un total de 340 euros
Cliente : Ah, je pensais que le total serait de 225 euros. Pouvez-vous vérifier, s'il vous plaît ?
Vendeur : oui effectivement, avez vous la carte de fidélité
Cliente : Non, je n'ai pas de carte de fidélité. Est-ce que ça vaut le coup d'en faire une ?
Vendeur : oui, elle ne coute que 10 euros, c'est une adhésion annuelle, grâce a votre achat du jour de 22 points, à 50 points vous auriez un bon d'achat de 5 euros
Cliente : D'accord, dans ce cas, je vais prendre la carte de fidélité aussi. Merci pour l'information !
Vendeur : ca fera un total de 235 euros, vous souhaitez regler comment ?
Cliente : Je vais régler par carte bancaire, s'il vous plaît.

Développer des services de formation s'adaptant aux conditions atmosphériques hyperlocales : une preuve de concept avec l'environnement FACTS

Imran S. A. Khan [0009-0000-2777-1663], Emmanuel G. Blanchard [0009-0006-0454-5739], Sébastien George [0000-0003-0812-0712]

Laboratoire d'Informatique de l'Université du Mans, Le Mans, FRANCE
{imran.khan,emmanuel.blanchard,sebastien.george}@univ-lemans.fr

Résumé : Les dérèglements climatiques font apparaître une nécessité d'adaptation pour les sociétés et les individus. Face à l'ampleur de la tâche et des enjeux, les environnements informatiques pour l'apprentissage humain pourraient contribuer à enseigner et diffuser de bonnes pratiques et connaissances facilitant ces adaptations nécessaires. Cet article présente une approche innovante d'application éducative dans laquelle des activités d'apprentissage ancrées dans un lieu sont directement adaptées en fonction des conditions atmosphériques. Une évaluation préliminaire est présentée et fournit des résultats très encourageants.

Mots-clés : IA générative, Internet des Objets, génération augmentée par récupération, adaptation atmosphérique, apprentissage ancré dans un lieu.

Abstract: Climate change has highlighted the need for societies and individuals to adapt. Given the scale of the task and the stakes involved, computer environments for human learning could be used to help teach and disseminate good practices and knowledge that facilitate these necessary adaptations. This article presents an innovative approach to educational applications in which place-based learning activities are directly adapted to atmospheric conditions. A preliminary evaluation is also presented, and prove to be very encouraging.

Keywords: Generative AI, Internet of Things, Retrieval Augmented Generation, Atmospheric Adaptation, Place-based Learning.

1 Introduction

Les changements climatiques accroissent la fréquence des événements météorologiques extrêmes tels que les vagues de chaleur, les sécheresses et les tempêtes, tout en en amplifiant leurs effets [10]. Ces perturbations affectent de nombreux secteurs de la société et de l'économie, qui dépendent fortement de conditions climatiques stables. Par exemple, dans le domaine de la viticulture, de nouvelles opportunités économiques émergent dans des régions qui, jusqu'à récemment, n'étaient pas propices à la culture de certaines variétés de raisin ou à la production de types spécifiques de vin. Parallèlement, les régions viticoles traditionnelles rencontrent des difficultés face à l'élévation des températures et à l'irrégularité des précipitations, ce qui impacte à la fois la quantité et la qualité de la production [11, 12]. Dans ce contexte, la diffusion des compétences et des connaissances en matière de résilience climatique devient de plus en plus essentielle pour assurer la prospérité future des individus, des industries et des sociétés [22].

Nous considérons que les technologies éducatives adaptatives peuvent jouer un rôle clé dans la transmission des savoirs et des compétences liés à la résilience climatique. C'est pourquoi nous proposons une nouvelle catégorie de technologies éducatives, fondée sur la combinaison d'éléments issus de l'Internet des objets (IoT) et de l'intelligence artificielle générative (IAGen), afin de concevoir des activités pédagogiques s'adaptant aux conditions atmosphériques.

Dans la suite de cet article, nous présentons un bref état de l'art, avant d'introduire notre technologie : le *Future Atmospheric Conditions Training System* (FACTS). Celle-ci est évaluée dans la section suivante. Les résultats sont discutés plus en profondeur dans la dernière partie ou sont également discutées des pistes de travail futures.

2 État de l'art

Cette brève revue de la littérature propose un aperçu des recherches et initiatives en matière d'éducation à la résilience climatique, ainsi que de l'intégration des technologies de l'IoT et de l'IAGen dans le domaine éducatif.

Éducation à la résilience climatique. Un consensus croissant se dégage quant à la nécessité de transformations profondes pour permettre aux individus et aux communautés de s'adapter aux effets du changement climatique [22, 31]. En conséquence, les efforts se multiplient pour développer des savoirs et bonnes pratiques ainsi que des politiques afin de renforcer la résilience climatique en milieu rural [9] aussi bien qu'en milieu urbain [6]. Dans ce contexte, les programmes de formation à la résilience climatique se développent, l'apprentissage ancré dans le territoire (*place-based learning*) apparaissant comme une démarche particulièrement appropriée [22]. Cette approche est définie comme « *une démarche pédagogique qui met l'accent sur le lien entre le processus d'apprentissage et le lieu physique où enseignants et apprenants se trouvent* » [34], les observations de terrain y étant reconnues comme particulièrement bénéfiques [13]. Pour que les savoirs et les bonnes pratiques en matière de résilience climatique soient efficaces, il est essentiel qu'ils soient diffusés de manière claire et convaincante auprès du grand public comme des professionnels, afin d'en favoriser une adoption

large [8]. Les Environnements Informatiques pour l'Apprentissage Humain (EIAH) peuvent faciliter cette communication [26]. Parmi les initiatives notables, on trouve l'utilisation de simulations éducatives pour simplifier la découverte des sciences du climat [8], le recours à la réalité virtuelle pour enseigner des concepts climatiques ou proposer des expériences immersives [15, 33], ainsi que la prise en compte des dimensions socioculturelles dans la conception des dispositifs EIAH destinés à la formation à la résilience climatique [1,26]. Enfin, bien que certaines initiatives existent qui utilisent les EIAH pour de l'éducation ancrée dans le territoire (par exemple [17, 39]), cette démarche est peu explorée dans le cadre de la résilience climatique.

IoT et éducation. L'Internet des objets est de plus en plus intégré dans des projets en éducation, tant pour améliorer la gestion des établissements et soutenir les enseignants que pour enrichir l'expérience d'apprentissage des élèves [36, 36]. Plusieurs initiatives exploitent des capteurs IoT pour personnaliser l'apprentissage, notamment en détectant les états affectifs des apprenants [28] ou en inférant leurs états mentaux [30]. Cependant, alors que les capteurs IoT sont couramment utilisés pour surveiller les conditions atmosphériques (température, qualité de l'air, etc.) [25], peu de projets les emploient en contexte éducatif [14]. De fait, les travaux intégrant des données atmosphériques au sein de systèmes d'apprentissage adaptatifs restent limités, suggérant une piste prometteuse pour de futures recherches alliant personnalisation de l'apprentissage et enjeux environnementaux.

IA générative en éducation. L'usage de l'intelligence artificielle générative (IAGen) en éducation connaît une croissance rapide et concerne des sujets variés tels que faciliter la correction de réponses courtes [23], générer des questions [7], faciliter la mise en place d'apprentissage par perturbation [5] ou aider à la conception de programmes pédagogiques [29]. Parmi les techniques émergentes, la génération augmentée par récupération (*Retrieval-Augmented Generation* – RAG) suscite un intérêt croissant [18]. La RAG permet notamment de réduire les risques d'hallucination [4] et se distingue par sa capacité à influencer la génération de contenu au moyen d'une base de connaissances, un mécanisme qui rappelle le module expert des systèmes tutoriels intelligents [2]. En combinant les modèles à base de récupération d'informations et les techniques génératives, la RAG ouvre des perspectives intéressantes pour la création d'expériences éducatives plus dynamiques, réactives et personnalisées. Cette approche est d'ores et déjà mise en œuvre dans diverses technologies éducatives intelligentes, par exemple pour renforcer l'engagement des apprenants [32], aider les tuteurs à adapter dynamiquement le style des systèmes d'enseignement [19], ou encore enseigner la cybersécurité à partir de connaissances structurées par des ontologies [38].

Les paragraphes précédents ont montré que l'IAGen et l'IoT apportent des contributions différentes mais complémentaires, à la fois pour intégrer les conditions atmosphériques et pour concevoir des dispositifs d'apprentissage adaptatifs. Pourtant, aucune recherche n'a encore exploré l'usage combiné de ces deux technologies pour créer des outils éducatifs adaptés aux variations climatiques. Nous faisons l'hypothèse qu'un tel système pourrait jouer un rôle important dans l'enseignement de la résilience climatique, en lien avec le contexte local. La suite de cet article explore l'hypothèse de recherche suivante :

Il est possible de combiner la IAGen et l'IoT pour concevoir des activités d'apprentissage ancrées dans le territoire, adaptables aux conditions atmosphériques en temps réel.

3 L'environnement FACTS (*Future Atmospheric Conditions Training System*)

FACTS est une technologie éducative adaptative intégrant des éléments d'IAGen et d'IoT. Elle propose des activités d'apprentissage ajustées en temps réel aux conditions atmosphériques locales. Fonctionnant comme un compagnon d'apprentissage, elle vise à former les utilisateurs à réagir à de nouvelles conditions climatiques, en s'inspirant des « savoirs » issus de régions ayant une plus grande familiarité avec ces conditions.

Dans cet article, nous nous appuyons sur un scénario de test dans lequel les utilisateurs incarnent un agriculteur en reconversion vers la viticulture. Ne venant pas d'une région traditionnellement viticole, ils ne disposent pas de repères sensoriels leur permettant de relier les conditions climatiques observées à des décisions viticoles appropriées. Les activités proposées par FACTS visent à les accompagner dans la construction de ces liens. Pour cela, la base de connaissances de FACTS a été initialisée avec une vingtaine d'articles scientifiques sur la viticulture en région méditerranéenne, mettant en évidence les relations entre conditions atmosphériques et pratiques culturales. La figure 1 présente l'architecture du système FACTS.

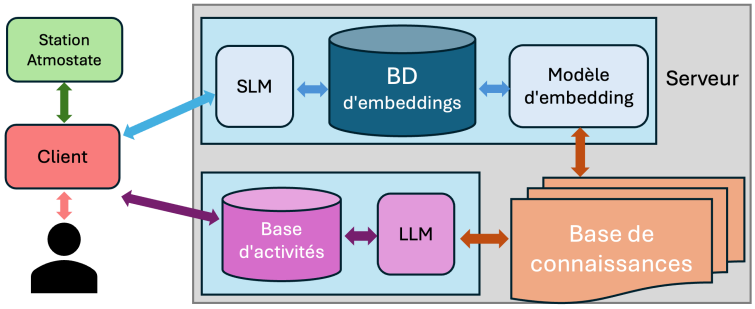


Fig. 1. L'architecture de l'environnement FACTS

Le système repose sur trois composants :

- Une application mobile, utilisée par l'apprenant pour réaliser les activités et interagir avec le système.
- Des stations Atmostate, appareils IoT qui mesurent les conditions atmosphériques locales.
- Une application serveur, qui utilise l'IA générative pour adapter les activités selon les données recueillies et la base de connaissances.

L'application client. La figure 2 présente des captures d'écran de l'application client.

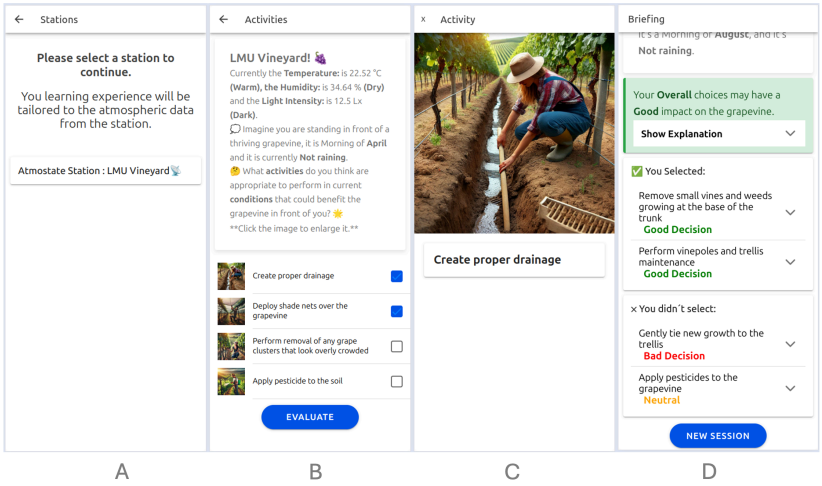


Fig. 2. Quatre captures d'écran de l'application client de FACTS

Une activité d'apprentissage avec FACTS se décompose en plusieurs étapes associées à des écrans de l'application client. Elle débute par une page simple constituée d'un message d'accueil et d'un unique bouton. Cliquer sur celui-ci permet d'accéder à une page présentant la liste des stations Atmostate détectées à proximité immédiate de l'utilisateur (fig. 2, A). L'utilisateur clique sur celle qui lui convient le mieux et fait ainsi apparaître l'écran présentant le défi. Celui-ci résume le contexte de l'activité (haut de fig. 2, B) qui est constitué des informations atmosphériques provenant des capteurs de la station Atmostate ainsi que d'informations additionnelles provenant d'API web géolocalisées tels que la vitesse du vent, conditions météo. Une liste d'actions est également proposée (haut de fig. 2, B) et cliquer sur l'une d'entre-elle permettra d'en afficher une illustration (fig. 2, C). Le défi pédagogique consiste à sélectionner lesquelles de ces actions sont appropriées en fonction des conditions atmosphériques. Les choix de l'utilisateur seront par la suite traités afin de générer un page de résultats commentant la pertinence des choix de l'utilisateur en fonction du contexte (fig. 2, D).

Les stations Atmostates. Spécifiquement conçues dans le cadre du projet FACTS, ces stations sont disséminées sur un territoire, dans des lieux sujets à des conditions atmosphériques distinctes. La figure 3 présente une station Atmostate ainsi qu'un exemple de répartition de 4 stations dans une zone de campagne mêlant haies de bocage, champs et forêt.

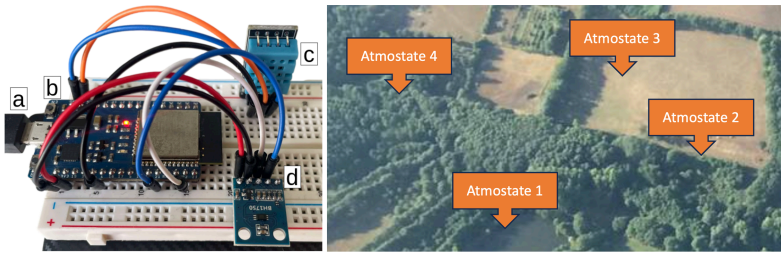


Fig. 3. Une station « Atmostate » et un exemple de répartition de plusieurs de ces stations

Une station Atmostate est alimentée par une pile 5 volts (Fig. 3, a) et est actuellement constituée d'un capteur de température et d'humidité DHT20 (Fig. 3, c), d'un capteur d'intensité lumineuse BH1750 (Fig. 3, d) et d'un microcontrôleur ESP32 (Fig. 3, b). Ce dernier a pour rôle de centraliser les données des différents capteurs et de les transmettre à l'application client via un signal WIFI. Le protocole MQTT est utilisé pour envoyer des paquets de données au format JSON toutes les 5 secondes, par exemple :

```
{"Temperature": "35°C", "Humidity":"75 %", "Light":"3000 Lx"}
```

Le serveur. Le serveur contient une base de connaissances exclusivement composée de documents textuels. Dans le cas de notre étude, elle a été constituée à partir d'une dizaine d'articles scientifiques discutant de moyens d'adapter l'agriculture et la viticulture au changement climatique. Cependant, des documents moins académiques pourraient tout aussi bien être considérés (ex : interviews de viticulteurs experts). Cette base de connaissances sert à nourrir deux processus à base d'IA générative.

- **Processus préparatoire**. Ce premier processus d'IA générative ne sera sollicité que lorsque la base de connaissances est créée ou modifiée. Un LLM analyse celle-ci pour en extraire les actions qui peuvent être appropriées ou non en fonction des conditions atmosphériques y est documentée. Dans notre scénario exemple, une vingtaine d'actions distinctes ont été identifiées et stockées dans une base de données d'actions. Celle-ci est interrogée à l'initialisation de tout nouveau défi pédagogique afin d'extraire aléatoirement 4 actions qui constitueront les choix proposés à l'utilisateur (bas de fig. 2, ii).
- **Processus de génération d'un défi atmosphériquement adapté**. Avant toute interaction, la base de connaissances textuelle est transformée en représentations numériques (appelées embeddings) à l'aide d'un modèle spécialisé. Lorsqu'un défi pédagogique est lancé, l'application crée un prompt combinant les données atmosphériques captées par une station Atmostate et les choix d'activités faits par l'utilisateur, supposés adaptés au climat du moment. Un processus RAG est alors activé : il enrichit le prompt en y ajoutant les embeddings les plus pertinents issus de la base de connaissances. Ce prompt enrichi est ensuite envoyé à un petit modèle de langage (SLM), qui génère une réponse claire et grammaticalement correcte, expliquant si les choix de l'utilisateur sont appropriés. Un exemple de prompt utilisé est présenté ci-dessous :

```
Using the {Retrieved Docs} as context, assess the suitability of
the selected {selected_activities} and unselected activities {unse-
lected_activities} under the given conditions: Temperature: 35°C
(Warm), Humidity: 75% (High), Light: 3000 lumens (Dark), Weather:
Cloudy, Rain: True, Period: Mid-April, Time: Evening. Evaluate each
activity as Good or Bad, with a brief justification. Also, provide
an overall assessment of whether the user's selection was appropri-
ate for a viticulture farm.
```

4 Évaluation

La RAG commence par rechercher des documents pertinents (retrieval) puis génère une réponse à partir des informations trouvées (*generation*). L'objectif est de produire une réponse fidèle aux sources et qui transmette correctement leur contenu. Pour évaluer cette capacité, on utilise souvent la métrique BERTScore F1, qui compare la réponse générée soit à une réponse de référence, soit directement aux documents sources. Cela permet de vérifier si le système comprend bien les documents consultés et s'il peut produire des réponses cohérentes, précises et conformes à l'information d'origine.

L'évaluation technique de notre système vise donc à évaluer la qualité de la RAG au moyen du BERTScore F1. Elle se décompose en 2 étapes :

- Phase de conversion en embeddings. Le texte de la base de connaissances est converti en *embeddings* en associant des modèles d'embeddings avec des méthodes de découpage (chunking). 3 modèles d'*embeddings* ont été testés : *nomic-embed-text, [21], bge-m3 [3] et snowflake-arctic-embed3 [27]*. Les 2 méthodes de découpage utilisées sont le *Recursive Chunking* [recur] qui segmente le texte de manière hiérarchique et l'*Agent-based Chunking* [agent], où les données sont introduites dans un agent (comme un SLM) qui génère les morceaux.
- Phase d'évaluation. On utilise différentes combinaisons de SLM et de modèles d'*embedding*s pour faire des générations. Les réponses générées sont ensuite évaluées en calculant le BERTScore. Dans ce cadre, nous avons extrait manuellement 32 paires de questions-réponses issues de la base de connaissances. Trois SLM ont été testés : *Llama3.2 [20], Phi3.5 [24] et Gemma:2b [16]*. Le tableau 1 présente les BERTScore F1 obtenus pour les différentes combinaisons.

Tableau 1. BERTScore F1 pour chaque combinaison de SLM et modèle d'embedding.

Modèles d'embedding	**SLM**		
	Llama3.2	Phi3.5	Gemma:2b
[recur] nomic-embed-text	0.89	0.89	0.88
[recur] bge-m3	0.88	0.89	0.88
[recur] snowflake-arctic-embed3	0.88	0.89	0.87
[agent] nomic-embed-text	0.88	*0.81*	0.87
[agent] bge-m3	0.88	0.89	0.88
[agent] snowflake-arctic-embed3	0.88	0.9	0.87

Par convention, un BERTScore F1 supérieur ou égal à 0,85 est considéré comme bon. Cela signifie que le texte généré a un sens très proche de celui du texte de référence, même si les mots ne sont pas identiques. Le tableau 1 montre que la combinaison du petit modèle de langage Phi3 avec le modèle d'embedding snowflake-arctic-embed2, utilisant la méthode agent-based chunking, a légèrement surpassé les autres. Toutefois, presque toutes les configurations testées ont dépassé ce seuil de qualité, sauf une. Ces résultats indiquent que les réponses générées par notre système, basé sur un petit modèle de langage et une architecture RAG, s'appuient bien sur les informations contenues dans la base de connaissances.

Nous avons aussi mené une première évaluation de l'expérience utilisateur auprès de 19 personnes (7 femmes et 12 hommes), avec une moyenne d'âge de 35 ans (écart-type : 12,56). Les retours ont été très positifs : les participants ont trouvé le système facile à utiliser et n'ont pas eu besoin d'aide pour comprendre ou résoudre les défis. Ils ont également apprécié l'originalité des activités, ainsi que l'usage de l'IA générative dans un cadre plus interactif et utile qu'un simple chatbot. Si ces premiers résultats sont encourageants, des évaluations à plus grande échelle permettront d'affiner ces conclusions.

5 Discussion et conclusion

Cet article présente FACTS, un système combinant des techniques d'IAGen, des capteurs IoT et une base de connaissances, afin de générer des défis pédagogiques adaptés aux conditions atmosphériques dans le cadre d'un apprentissage ancré dans un lieu, portant sur la résilience climatique.

Une évaluation de plusieurs combinaisons de SLM et modèles d'*embedding* a confirmé que l'approche RAG au cœur de FACTS fonctionnait adéquatement et généraient des réponses à des défis adaptées aux conditions atmosphériques qui provenaient directement d'informations de notre base de connaissance. Une évaluation préliminaire a mis en évidence que les activités proposées par FACTS étaient bien comprises des utilisateurs, qui suggéraient d'ailleurs d'autres domaines d'utilisation.

L'un des principaux atouts de FACTS est justement son usage de l'approche RAG qui permet de générer facilement des défis pédagogiques adaptés aux conditions atmosphériques, dans de nombreux domaines, simplement en mettant à jour la base de connaissances. Cependant, construire une base de connaissances vraiment pertinente reste un défi, notamment pour identifier des ressources qui relient de manière claire des activités d'apprentissage à des situations atmosphériques spécifiques. Dans cette étude, des articles scientifiques ont été utilisés, mais des sources alternatives, comme des entretiens avec des experts locaux, pourraient enrichir le système avec des connaissances plus concrètes, parfois absentes de la littérature académique.

La version actuelle de FACTS se concentre sur des scénarios liés à la viticulture. Toutefois, les participants à notre évaluation ont souligné que l'outil pourrait aussi être utile dans d'autres secteurs comme l'agriculture, la biologie marine, l'apiculture ou la foresterie. Dans les travaux à venir, nous prévoyons d'élargir nos tests à des groupes

plus importants et à ces domaines, qui partagent tous le besoin d'une formation adaptée à la résilience climatique.

References

1. Aguayo, C., & Eames, C.: Promoting community socio-ecological sustainability through technology: A case study from Chile. International Review of Education, 63, 871-895 (2017)
2. Anderson, J. R., Boyle, C. F., & Reiser, B. J. (1985). Intelligent tutoring systems. Science, 228(4698), 456-462.
3. BGE-M3, Embedding model: https://huggingface.co/BAAI/bge-m3 . Last accessed 2025/05/10.
4. Béchard, P., & Ayala, O. M.: Reducing hallucination in structured outputs via Retrieval-Augmented Generation. arXiv preprint (2024). arXiv:2404.08189.
5. Blanchard, E.G, Callahan, J-C, Akretche, I., Asswiel, N., Schmitt, L., Kessemtini, A.,Khan, I.S.A.: Making Generative AI Hallucinations Useful by Reassessing the Troublemaker Agent Strategy. To appear in Late Breaking Results of the 26th International Conference on Artificial Intelligence in Education (2025).
6. Brown, C., Shaker, R. R., & Das, R.: A review of approaches for monitoring and evaluation of urban climate resilience initiatives. Environment, development and sustainability, 20, 23-40 (2018).
7. Bulathwela, S., Muse, H., Yılmaz, E.: Scalable Educational Question Generation with Pre-trained Language Models. In: Artificial Intelligence in Education. pp 327–339 (2023).
8. Bush, D., Sieber, R., Seiler, G., & Chandler, M.: Examining educational climate change technology: how group inquiry work with realistic scientific technology alters classroom learning. Journal of Science Education and Technology, 27, 147-164 (2018)
9. Chirisa, I., & Nel, V.: Resilience and climate change in rural areas: a review of infrastructure policies across global regions. Sustainable and Resilient Infrastructure, 7(5), 380-390 (2022)
10. Clarke, B., Otto, F., Stuart-Smith, R., Harrington, L.: Extreme weather impacts of climate change: an attribution perspective. Environ. Res.: Climate. 1, 012001 (2022). https://doi.org/10.1088/2752-5295/ac6e7d.
11. Droulia, F., & Charalampopoulos, I.: Future climate change impacts on European viticulture: A review on recent scientific advances. Atmosphere, 12(4), 495 (2021).
12. Fraga, H.: Viticulture and Winemaking under Climate Change. Agronomy. 9, 783 (2019). https://doi.org/10.3390/agronomy9120783.
13. Haskin, J.: Place-based learning: The technology frontier in environmental education. *Educational Technology*, *39*(6), 59-63 (1999).
14. Hu, Y., Huang, R.: Development of Weather Monitoring System Based on Raspberry Pi for Technology Rich Classroom. In: Chen, G., Kumar, V., Kinshuk, Huang, R., and Kong, S.C. (eds.) Emerging Issues in Smart Learning. pp. 123–129. Springer Berlin Heidelberg, Berlin, Heidelberg (2015).
15. Klippel, A., Zhao, J., Jackson, K. L., La Femina, P., Stubbs, C., Wetzel, R., & Oprean, D.: Transforming earth science education through immersive experiences: Delivering on a long held promise. Journal of Educational Computing Research, 57(7), 1745-1771 (2019).
16. Gemma 2B, Language Model: https://huggingface.co/google/gemma-2b, last accessed 2025-05-10
17. Levy, D., Shafriri, Y., & Alef, Y.: A place to discover, imagine, and change: Smart learning with local places. In Universal Access in Human-Computer Interaction. Theory, Methods and Tools: 13th International Conference (2019).

18. Lewis, P.S.H., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.-T., Rocktäschel, T., Riedel, S., Kiela, D.: Retrieval-Augmented Generation for Knowledge-Intensive NLP tasks. Neural Information Processing Systems. 33, 9459–9474 (2020).
19. Liu, C., Hoang, L., Stolman, A., Wu, B.: HiTA: A RAG-Based Educational Platform that Centers Educators in the Instructional Loop. In: Olney, A.M., Chounta, IA., Liu, Z., Santos, O.C., Bittencourt, I.I. (eds) Artificial Intelligence in Education. AIED 2024., vol 14830. Springer, Cham (2024).
20. Llama3.2, Language Model: https://huggingface.co/meta-llama/Llama-3.2-3B . Last accessed 2025/02/19.
21. Nomic-Embed-Text, Embedding model: https://huggingface.co/nomic-ai/nomic-embed-text-v1 . Last accessed 2025/05/10.
22. Nusche, D., Rabella, M. F., & Lauterbach, S. (2024). Rethinking education in the context of climate change: Leverage points for transformative change. OECD education working papers, 307, 1-71 (2024).
23. Nye, B. D., Mee, D., Core, M.G.: Generative Large Language Models for Dialog-Based Tutoring: An Early Consideration of Opportunities and Concerns. In Proc. of the AIED Workshop on Empowering Education with LLMs - the Next-Gen Interface and Content Generation, pp 78-88 (2023).
24. Phi3.5, Language Model: https://huggingface.co/microsoft/Phi-3.5-mini-instruct, last accessed 2025/02/19.
25. Salam, A.: Internet of things for environmental sustainability and climate change. In Internet of Things for sustainable community development: Wireless communications, sensing, and systems (pp. 33-69). Cham: Springer International Publishing (2024).
26. Sidze, S. M.: Empowering Community Resilience to climate Change in Cameroon using Technology-enhanced Learning. PhD Thesis, Dresden University of Technology (2017). https://core.ac.uk/download/pdf/236374643.pdf . Last accessed April 10, 2025
27. Snowflake-Arctic-Embed, Embedding moded: https://github.com/Snowflake-Labs/arctic-embed. Last accessed 2025/05/10
28. Spyrou, E., Vretos, N., Pomazanskyi, A., Asteriadis, S., & Leligou, H. C.: Exploiting IoT technologies for personalized learning. In Proc. of IEEE Conference on Computational Intelligence and Games, 1-8 (2018).
29. Sridhar, P. Doyle, A., Agarwal, A. Bogart, C., Savelka, J., Sakr, M. Harnessing LLMs in curricular design. Using GPT-4 to support authoring of learning objectives. In Proc. of the AIED Workshop on Empowering Education (2023).
30. Taherisadr, M., Al Faruque, M. A., & Elmalaki, S.: Erudite: Human-in-the-loop IoT for an adaptive personalized learning system. IEEE Internet of Things Journal, 11(8), 14532-14550 (2023)
31. Tammi, A. M., & Martínez, R.: Transforming Education Systems for a Climate-Resilient and Sustainable Future. Building Resilience in Education Systems, 142 (2025).
32. Thüs, D., Malone, S., Brünken, R.: Exploring generative AI in higher education: a RAG system to enhance student engagement with scientific literature. Front. Psychol. 15, 1474892 (2024)
33. Tisoglu, S., Sönmez, E., Kaya, K. Y., & Et, S. Z.: Bridging pedagogy and technology: a systematic review of immersive virtual reality's potential in climate change education. Environmental Education Research, 1-31 (2025)
34. Yemini, M., Engel, L., & Ben Simon, A.: Place-based education–a systematic review of literature. Educational Review, 77(2), 640-660 (2025).

35. Zeeshan, K., Hämäläinen, T., & Neittaanmäki, P.: Internet of Things for sustainable smart education: An overview. Sustainability, 14(7), 4293 (2022).
36. Zhang, X., Ding, Y., Huang, X., Li, W., Long, L., Ding, S.: Smart Classrooms: How Sensors and AI Are Shaping Educational Paradigms. Sensors. 24, 5487 (2024).
37. Zhang, T., Kishore, V., Wu, F., Weinberger, K. Q., & Artzi, Y.: Bertscore: Evaluating text generation with bert. *arXiv preprint arXiv:1904.09675* (2019).
38. Zhao, C., Agrawal, G., Kumarage, T., Tan, Z., Deng, Y., Chen, Y. C., & Liu, H.: Ontology-Aware RAG for Improved Question-Answering in Cybersecurity Education. (2024). arXiv preprint arXiv:2412.14191. Last accessed April 10, 2025.
39. Zimmerman, H. T., & Land, S. M.: Facilitating place-based learning in outdoor informal environments with mobile computers. *TechTrends*, *58*, 77-83 (2014).

Tester son programme lors d'activités de conception de jeux vidéo à l'école

Jérémy Le Du[1], Julian Alvarez[2], Daniel Schmitt[1]

[1] Université de Lorraine
Jeremy.Le-Du@hrsk.fi
daniel.schmitt@univ-lorraine.fr
[2] Université de Lille
Julian.alvarez@univ-lille.fr

Résumé. Cette étude explore le rôle du test de programme dans le cadre d'activités de conception de jeux vidéo en milieu scolaire. Elle a été conduite auprès de 373 élèves âgés de 9 à 13 ans, dans le cadre du projet Educational Game & Play Design Project (EGPDP). S'appuyant sur une approche mixte combinant l'analyse de vidéos avec suivi oculaire et des entretiens en rappel stimulé (méthode Remind), ce travail montre qu'une majorité des élèves utilisent des tests pour valider un élément du jeu ou du programme nouvellement ajouté, tandis qu'une minorité les mobilise pour déboguer ou se divertir. Une forte corrélation positive (96 %) est observée entre la fréquence des tests et les scores post-test des élèves en matière de concepts de pensée informatique. Ces résultats semblent conforter la valeur pédagogique des pratiques de test et de débogage systématiques dans les activités de conception de jeux vidéo.

Mots-clés : pensée informatique, conception de jeux vidéo, test de progrmame, technologie éducative, recherche mixte, suivi oculaire, méthode Remind.

Abstract. This study explores the role of program testing within the framework of video game design activities in a school setting. It was conducted with 373 students aged 9 to 13 as part of the Educational Game & Play Design Project (EGPDP). Using a mixed-methods approach that combines video analysis with eye tracking and stimulated recall interviews (Remind method), this study shows that the majority of students use testing primarily to validate newly added game or program elements, while a minority use them for debugging or entertainment. A strong positive correlation (96%) was observed between test frequency and students' post-test scores in computational thinking concepts. These findings seem to support the educational value of systematic testing and debugging practices in video game design activities.

Keywords: Computational Thinking, Video Game Design, Program Testing, Educational Technology, Mixed-Methods Research, Eye Tracking, Remind Method.

1 Introduction

Aujourd'hui encore, la pensée informatique ne fait pas l'objet d'une définition consensuelle. Certains la considèrent comme une compétence du XXIe siècle en tant que telle [1], d'autres comme un processus de réflexion [2], un ensemble de compétences [3] ou un processus de résolution de problèmes [4], multidisciplinaire donc fondamental pour chaque enfant [5].

Un moyen d'opérationnaliser la pensée informatique est proposé par Brennan et Resnick [6], qui distinguent les concepts, les pratiques et les perspectives de la pensée informatique. Cet article se focalise sur l'une de ces pratiques, le test et le débogage, largement liée à d'autres puisqu'elle implique une réflexion logique et analytique pour isoler le problème et se concentrer sur l'erreur, et qu'elle fait partie intégrante des stratégies de développement incrémental et de décomposition des problèmes [7].

Une métasynthèse portant sur 55 études empiriques [8] montre que les tests et le débogage constituent la pratique informatique la plus fréquemment observée dans le cadre d'activités de programmation avec le logiciel Scratch. Les élèves semblent pratiquer les tests et le débogage de manière récurrente [9], à mesure que les programmes gagnent en complexité. Cette pratique se développe à un rythme lent, en particulier la capacité à déboguer un programme, processus complexe qui nécessite de la persévérance et une approche systématique [10], et qui se manifeste selon différentes stratégies [11]. Elle semble favorisée par la collaboration entre élèves [12], et contribue à l'amélioration de la capacité à analyser le code [10] et à comprendre le fonctionnement d'un programme [13].

Cette pratique informatique intègre l'identification de la source du problème, la lecture et la réécriture des scripts, la recherche d'exemples de scripts fonctionnels, toutes associées au débogage ; l'expérimentation des scripts, associée au test [6]. Les études consacrées demeurent rares dans des environnements de programmation par blocs [14; 11], voire inexistantes s'agissant spécifiquement de la pratique de test sur laquelle se focalise la présente étude. Celle-ci examine la capacité d'élèves de 9 à 13 ans à tester le programme informatique conçu. Les questions de recherche sont les suivantes :

- QR1 : À quelle fréquence les élèves testent-ils leur programme lorsqu'ils conçoivent un jeu vidéo ?
- QR2 : Quelles sont les attentes des élèves lorsqu'ils testent le programme créé ?
- QR3 : La pratique de test rend-elle compte de la maîtrise des concepts de la pensée informatique ?

2 Méthode

Une approche mixte, fondée sur le paradigme du pragmatisme [15] et qui combine les approches quantitatives et qualitatives pour se concentrer sur le produit final de la recherche [16], est suivie.

2.1 Contexte et participants

Cette étude est menée dans le cadre du projet Educational Game & Play Design (EGPDP), un programme de conception de jeux vidéo conçu pour être adapté à des élèves du CE2 à la 5e, et proposé à des établissements scolaires francophones de Finlande et de France depuis l'année scolaire 2022-23. Au total, 373 élèves sont inclus dans cette étude.

2.2 Educational Game & Play Design Project

La figure 1 résume les principales caractéristiques du programme EGPDP, définies à partir d'une étude pilote et d'une analyse critique des expériences antérieures de conception de jeux vidéo [17].

Les élèves conçoivent des jeux vidéo éducatifs destinés à leurs pairs avec le logiciel Scratch (6 sur figure 1). Ils programment par pair (4), dispositif qui a montré son efficacité [18]. Quinze séances de 45 minutes sont organisées (2), dont deux dédiées aux évaluations pré-test et post-test (9). Le module est divisé en deux phases différentes (1).

La première est consacrée à l'apprentissage de la programmation et des stratégies de conception de jeux. Elles s'appuient sur 10 modèles de jeu de difficulté variable (8). Les séances sont divisées en temps d'apprentissage de quinze minutes, organisées autour d'une vidéo tutoriel de trois minutes : des vidéos courtes, limitées à un seul objectif, basées sur des exemples concrets, suivies d'une mise en œuvre immédiate [19].

La seconde phase laisse place à la conception libre du jeu vidéo, un temps jugé suffisant pour stimuler la créativité des élèves en s'éloignant des modèles étudiés durant les séances précédentes. À deux reprises, des sessions de tests sont organisées.

Les enseignants disposent d'un ensemble de ressources pédagogiques à leur disposition : des séances de formation en ligne (5, 7) et divers outils pédagogiques. Leur rôle principal est de guider les élèves dans le processus de développement des jeux et dans l'appropriation des ressources.

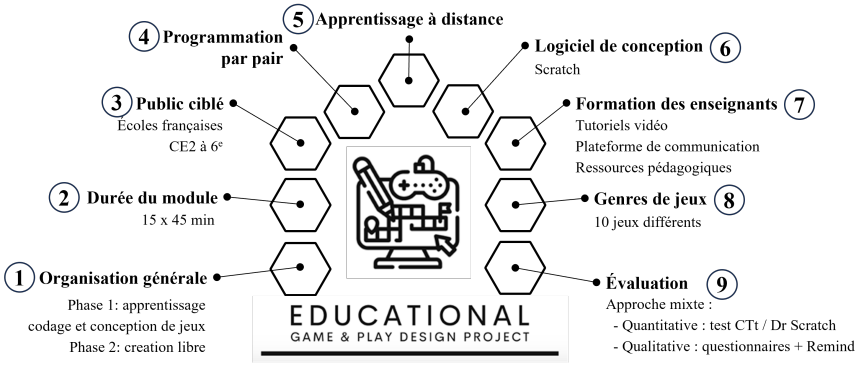


Fig. 1. Protocole expérimental suivi dans le EGPDP [17]

2.3 Mesure

Questionnaire d'auto-évaluation
Le premier outil de mesure repose sur une question posée aux élèves lors du post-test, leur demandant à quelle fréquence ils testent chaque nouvel élément ajouté à leur programme. Une échelle de Likert à 4 points est utilisée : Aucun ; Un peu ; Beaucoup ; Énormément.

Analyse de vidéos des séances
18 groupes d'élèves de niveau hétérogène en programmation ont été observés à l'aide de lunettes caméras avec suivi oculaire lors de la phase de conception. Les vidéos des séances ont ensuite été analysées afin d'identifier et mesurer la durée des phases de test, facilement discernables dans Scratch.

Méthode Remind
En complément, des entretiens en rappel stimulé avec perspective subjective située de 20/25 minutes ont été menés auprès des élèves précédents en suivant le protocole Remind [20]. La combinaison du suivi oculaire avec des entretiens en rappel stimulé caractérisant cette approche permet de comprendre une part significative du raisonnement et des processus de pensée complexes [21], en l'occurrence ici comment l'élève utilise la pratique de test.

Les entretiens ont ensuite été analysés à partir d'un langage de description basé sur l'utilisation de signes hexadiques permettant de raconter la dynamique d'une expérience vécue [22], couplé au modèle #5c21 [23] qui décompose cinq compétences du 21e siècle en composantes mesurables [24]. Dans le cadre de cette étude, seule la composante « s'engager dans le processus d'évaluation et d'amélioration itérative d'un programme informatique » associée à la pensée informatique est considérée.

Mesure des concepts de pensée informatique
L'évaluation des concepts de la pensée informatique s'est faite à l'aide d'une méthode quantitative, sous la forme d'un pré-test et d'un post-test appariés. Le CTt [25] ciblant les élèves du CM2 à la 3e, a démontré à la fois une validité de critère [26], une validité interculturelle [27], tout en disposant de la variante cCTt [28] adaptée aux classes de CE2/CM1. Ces deux tests fournissent une évaluation décontextualisée basée sur respectivement 28 et 25 questions à choix multiples. Dans cette étude, toutes les notes sont ramenées en pourcentage.

3 Résultats

3.1 Fréquence de test et développement des concepts informatiques

La figure 2 représente la distribution des élèves en fonction de leur fréquence de test. Pour mesurer une éventuelle différence dans les bénéfices des activités de conception

de jeux vidéo avec la fréquence de test, les scores moyens pré- et post-test des quatre groupes d'élèves considérés sont représentés. Une corrélation de 96 % est observée entre la fréquence de test et la note moyenne obtenue au post-test, corrélation indépendante de l'âge des élèves.

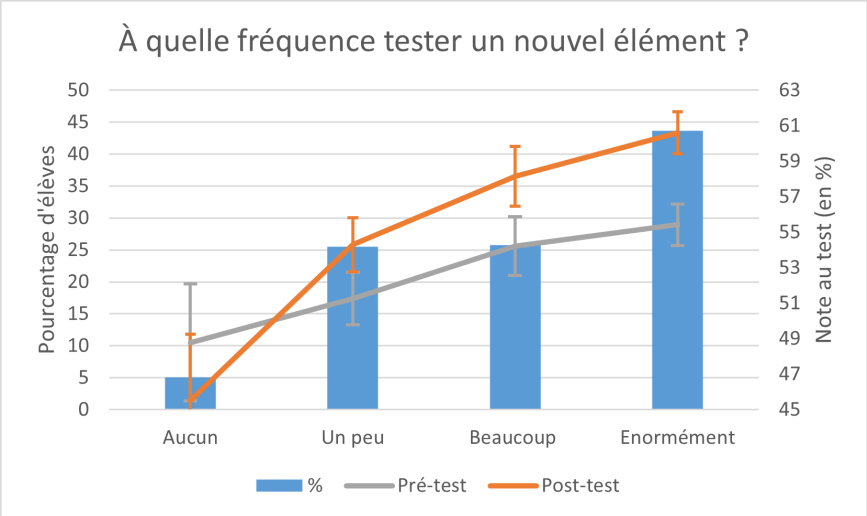


Fig. 2. Distribution des élèves en fonction de la fréquence de test (bleu), et score moyen obtenu au pré-test (gris) et au post-test (orange).

La figure 3 montre la variation de la fréquence de test en fonction de l'âge des élèves.

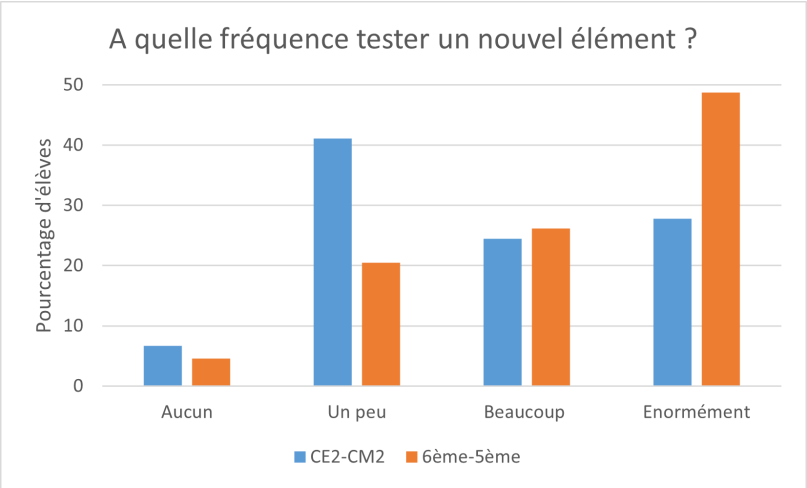


Fig. 3. Distribution des 90 élèves de CE2-CM2 (bleu) et 283 élèves de 6ème-5ème (orange) en fonction de la fréquence de test.

3.2 Analyse des séances

Analyse de vidéos des séances
L'analyse des vidéos en vue subjective des 18 groupes montre qu'en moyenne, les élèves consacrent 12,2 % du temps d'une séance à tester leur programme. Cela représente 11,6 séquences différentes pendant la séance, soit une moyenne de 30,2 secondes consacrée à chaque test, pour un écart type de 15,1s.

Les entretiens Remind des 18 groupes montrent qu'en moyenne, les élèves consacrent 6,9 % du temps de l'entretien à expliciter des actions liées au test de leur programme. Cela représente 5,7 moments différents pendant l'entretien, soit une moyenne de 11,4 secondes consacrée à chacun.

Concernant l'utilisation des tests faite par les élèves, sur les 97 verbalisations des élèves indiquant qu'ils sont en train de tester leur programme, 69 le font pour vérifier si le nouvel élément ajouté fonctionne conformément à leurs attentes, 13 mentionnent explicitement avoir découvert un nouveau problème suite à ce test, 14 testent spécifiquement le jeu en vue de résoudre un problème de codage qu'ils sont en train de traiter, et 1 élève indique tester le jeu simplement pour se divertir.

Fréquence de test et développement des concepts informatiques
La figure 4 représente les notes des élèves obtenues au pré-test et au post-test en fonction du pourcentage de temps consacré à tester le programme, estimé à partir de l'analyse des vidéos.

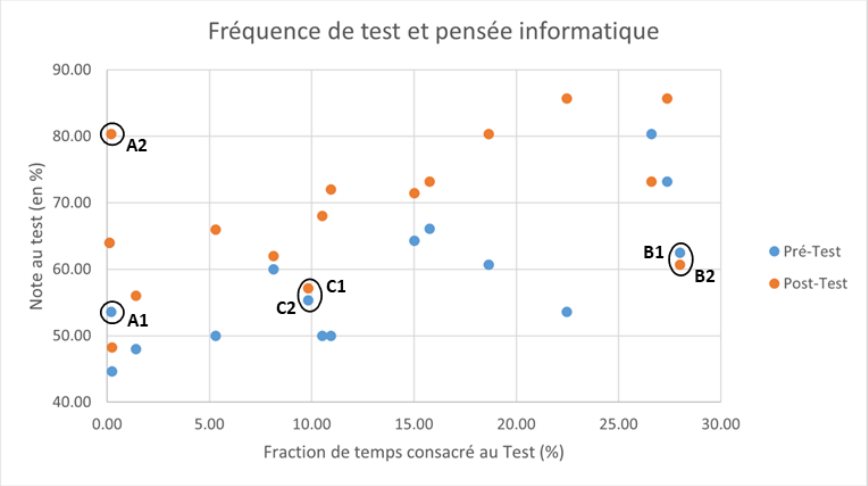


Fig. 4. Notes des élèves obtenues au pré-test et au post-test en fonction du pourcentage de temps de test.

4 Discussion

4.1 Temps de test du programme

La pratique de « tester » le programme conçu est largement mobilisée par les apprentis concepteurs. Lorsqu'ils ajoutent un nouvel élément à leur programme, ils sont une large majorité à le tester systématiquement ou presque, ce qui confirme des études antérieures [8, 9]. Seule une très faible proportion ne teste jamais le jeu conçu.

Ce résultat s'explique par la complexité de la tâche de conception d'un jeu vidéo, qui oblige à décomposer le jeu pensé en différents éléments à programmer, et pousse ainsi à adopter une stratégie de développement incrémental [7]. Il vient aussi confirmer l'efficacité de la phase 1 du programme, qui a su communiquer l'importance de cette pratique auprès d'une large majorité d'élèves novices en programmation. Plusieurs raisons semblent expliquer les 19 élèves n'ayant jamais testé leur programme : manque de motivation, forte confiance en leur capacité à programmer sans aide, forte imitation des jeux modèles.

L'analyse des vidéos en vue subjective confirme l'enquête menée auprès des 373 participants puisque les élèves consacrent en moyenne 12,2 % du temps à tester leur programme. Un constat similaire est opéré avec les entretiens Remind, la moindre fraction du temps consacré s'expliquant par le fait que cette pratique ne nécessite guère d'explication pour être comprise. Si la durée des entretiens avait été fixée à celle des séances, les écarts auraient sans doute été inexistants.

À partir de l'analyse des vidéos et les entretiens Remind, nous expliquons les écarts constatés entre les groupes sur le nombre total de tests opérés et le temps moyen d'un test de trois façons : l'âge des élèves, la séance observée et le type de jeu programmé.

L'âge des élèves impacte la propension des élèves à tester leur jeu. La figure 3 indique que les élèves les plus âgés testent davantage leur jeu. Les élèves de CE2 ont davantage tendance à s'inspirer du programme des jeux modèles, avant de le modifier et se l'approprier une fois leur propre jeu fonctionnel. Le test de chaque nouveau bloc ajouté ne constitue alors plus un impératif puisqu'ils basent leur « création » sur un programme fonctionnel.

La séance pendant laquelle les élèves sont observés est le second facteur expliquant les différences entre les groupes. Les 18 groupes ont été observés à différents niveaux d'avancement de leur jeu. Ainsi, les élèves du groupe 10 (A1/A2 sur la figure 4) en étaient à la séance 10, dans des phases créatives intenses, consacrant peu de temps à programmer des scripts. L'absence de test opéré n'implique pas qu'ils ne testent jamais leur programme, mais simplement qu'ils n'ont pas eu besoin de le faire durant cette séance.

Le type de jeu programmé impacte largement la durée d'un test associé. À titre d'exemple, les élèves du groupe 17 (C1/C2 sur la figure 4), qui ont créé un jeu de plateforme, ont multiplié les tests rapides pendant la séance, au contraire des élèves du groupe 1 (B1/B2 sur la figure 4) qui ont programmé un jeu narratif. Doit-on pour autant favoriser des « jeux courts » qui diminuent les temps de test ? Ce serait négliger l'aspect motivationnel, un facteur important pour justifier le potentiel éducatif des activités de

conception de jeux vidéo à l'école [29]. Les filles seraient ainsi plus motivées lorsque la programmation est intégrée à une activité narrative [30].

Plutôt que d'éliminer certains types de jeux, un renforcement des stratégies de test dans la première phase du programme EGPDP semblerait davantage profitable. La pratique de test et de débogage devrait faire l'objet d'un enseignement spécifique qui pourrait s'inspirer de certaines initiatives récentes [31 ; 32] et faire l'objet d'activités de débogage tangibles [33] à partir d'exemples concrets [34] et de mises en situation liées à la programmation de jeux vidéo.

4.2 Le rôle du test

Les entretiens Remind permettent de comprendre les raisons qui poussent les élèves à tester le jeu vidéo qu'ils programment.

Dans de nombreux cas relevés, test et débogage forment des pratiques associées. Lorsque les élèves testent leur programme, cela conduit parfois à la découverte d'un nouveau problème, qui les amène à tenter de le résoudre, puis de tester la correction apportée. Dans ce cas, le test évalue la solution proposée au problème de codage identifié et traité.

Néanmoins, cette entrée dans un cycle de test-débogage jusqu'à résolution du problème ne semble pas refléter la majorité. Celle-ci voit les tests plus souvent utilisés pour s'assurer du bon fonctionnement des nouveaux éléments ajoutés à leur jeu, comme de leur intérêt et leur difficulté pour le jeu lui-même. On se situe ici plus dans une dimension ludique qu'informatique, la création du jeu étant étroitement liée à la pratique du jeu [9]. Le test prend une dimension plus générale et intègre pleinement l'évaluation et l'amélioration itérative du jeu en tant que tel. Il fait office de validation pour l'élève, que celle-ci porte sur le code ou la conception du jeu, et s'accompagne souvent d'une émotion positive ressentie, voire d'une frustration lorsque le résultat ne correspond pas aux attentes.

Le dernier cas de figure rencontré concerne l'utilisation du test dans la seule optique de divertissement. L'échantillon d'élèves observé nous suggère que cette pratique est mineure. Le jumelage des élèves avec leurs amis semble lié à une interaction hors-jeu les éloignant de la tâche selon certaines études [18]. Nous n'observons pas cet effet à travers une hypothétique multiplication des tests à visée purement ludique.

Les tests sont ainsi le plus souvent utilisés pour évaluer le nouvel élément de jeu programmé, notamment son aspect ludique. En cela, les séances de test collectives offrent des retours d'expérience contribuant à l'équilibrage du jeu. Nous recommandons leur instauration lors de la mise en place d'activités de ce type à l'école.

4.3 Lien entre pratique de test et concepts informatiques

Les questionnaires d'auto-évaluation comme l'analyse des 18 groupes d'élèves observés montrent que les élèves qui testent le plus leur programme sont ceux qui obtiennent les meilleurs résultats au post-test, mais aussi ceux qui progressent le plus. L'abondance de tests opérés ne témoignerait donc pas d'un manque de compétences en programmation, amenant ces élèves à procéder par essais-erreurs pour corriger des bugs

par manque d'approche stratégique, niveau élémentaire des stratégies de débogage [11].

En rentrant dans une démarche de résolution de problèmes de codage et de développement incrémental, les phases de test constituent un maillon essentiel du prototypage itératif de création d'une solution à une situation-problème, associée ici à un élément du jeu vidéo conçu [35]. Les élèves ne l'ayant pas assimilé sont ainsi moins confrontés à des situations leur permettant d'assimiler les concepts nécessaires à la résolution d'un programme, encore moins à progresser dans un test mesurant leur maîtrise dans des situations décontextualisées. Au contraire, les élèves les plus à l'aise, pourtant les plus à même de se passer d'une phase validant l'ajout de scripts élémentaires, ont tendance à systématiquement tester chaque bloc, une pratique qui tend à se réduire à mesure qu'ils améliorent leur capacité à simuler mentalement l'exécution d'un programme et à utiliser des algorithmes [11].

Nous voyons ici comment concepts et pratiques informatiques sont liées. Il est donc essentiel que les programmes intègrent ces deux aspects dans leur plan d'apprentissage. La conception de jeux vidéo constitue un bon moyen le permettant, en particulier par rapport à des exercices de codage plus classiques [36], puisque les élèves sont contraints d'apprendre à décomposer leur jeu en différents éléments, à les programmer de manière itérative en faisant preuve d'adaptation, à isoler un problème pour le tester et le déboguer, à réutiliser des parties de scripts fonctionnels issus d'autres jeux, autant de pratiques renforçant la pensée informatique de l'élève selon toutes ses dimensions.

5 Conclusion

L'objectif de cet article est d'examiner comment, lors d'activités de conception de jeux vidéo, se manifeste la pratique de test. L'étude menée auprès de 373 élèves de 9 à 13 ans ayant participé au programme EGPDP nous permet de mesurer l'importance de cette composante de la pensée informatique, largement mobilisée par les participants.

Le test est majoritairement utilisé par les élèves pour évaluer un élément du jeu ou du programme, et prend une place essentielle dans le processus d'évaluation et d'amélioration itérative du programme créé. Dans certains cas, minoritaires, le test s'accompagne de la découverte d'un problème et de l'entrée dans un cycle de test-débogage. Le test s'intègre alors pleinement dans le processus de résolution du problème puisqu'il valide ou invalide la solution proposée.

La fréquence de test est positivement corrélée aux résultats du test mesurant la maîtrise des concepts de la pensée informatique. Cela renforce l'idée d'un développement conjoint des concepts et des pratiques de la pensée informatique dans les programmes d'enseignement. Ces dernières ne sont pas naturellement maîtrisées par les apprenants et nécessitent un apprentissage spécifique.

Cette étude pourrait être développée. Un échantillon plus important d'élèves nous permettrait de davantage comprendre les différentes utilisations du test en fonction du profil des élèves, et comment le type de jeu conçu impacte celles-ci. L'influence de l'enseignant dans le développement de la pratique de test pourrait aussi être étudiée.

References

1. Voogt, Joke, Petra Fisser, Jon Good, Punya Mishra, et Aman Yadav. 2015. « Computational Thinking in Compulsory Education: Towards an Agenda for Research and Practice ». *Education and Information Technologies* 20(4):715-28. doi: 10.1007/s10639-015-9412-6.
2. Wing, JM. 2011. « A definition of computational thinking from Jeannette Wing ». *Computing educational research blog*.
3. Fagerlund, Janne, Päivi Häkkinen, Mikko Vesisenaho, et Jouni Viiri. 2021. « Computational Thinking in Programming with Scratch in Primary Schools: A Systematic Review ». *Computer Applications in Engineering Education* 29(1):12-28. doi: 10.1002/cae.22255.
4. Standl, Bernhard. 2017. « Solving Everyday Challenges in a Computational Way of Thinking ». P. 180-91 in *Informatics in Schools: Focus on Learning Programming*. Vol. 10696, édité par V. Dagienė et A. Hellas. Cham: Springer International Publishing.
5. Denning, Peter. 2017. « Computational Thinking in Science ». *American Scientist* 105:13-17.
6. Brennan, K., et M. Resnick. 2012. « New Frameworks for Studying and Assessing the Development of Computational Thinking ». P. 25 in *2012 Annual Meeting of the American Educational Research Association*. Vol. 1. Vancouver.
7. Grover, Shuchi, et Roy Pea. 2017. « Computational Thinking: A Competency Whose Time Has Come ». P. 20-38 in *Computer Science Education: Perspectives on Teaching and Learning in School*. Bloomsbury Publishing.
8. Zhang, LeChen, et Jalal Nouri. 2019. « A systematic review of learning computational thinking through Scratch in K-9 ». *Computers & Education* 141:103607. doi: https://doi.org/10.1016/j.compedu.2019.103607.
9. Kafai, Yasmin, et Veena Vasudevan. 2015. « Constructionist Gaming Beyond the Screen: Middle School Students' Crafting and Computing of Touchpads, Board Games, and Controllers ». P. 49-54 in.
10. Falloon, G. 2016. « An analysis of young students' thinking when completing basic coding tasks using Scratch Jnr. On the iPad: General thinking and computational work ». *Journal of Computer Assisted Learning* 32. doi: 10.1111/jcal.12155.
11. Yan, Wei, Feiya Luo, Maya Israel, Ruohan Liu, et Latoya Chandler. 2025. « How Do Elementary Students Apply Debugging Strategies in a Block-Based Programming Environment? » *Education Sciences* 15:292. doi: 10.3390/educsci15030292.
12. Cafarella, Lorien, et Lucas Vasconcelos. 2024. « Computational thinking with game design: An action research study with middle school students ». *Education and Information Technologies* 30:5589-5633. doi: 10.1007/s10639-024-13010-5.
13. Webb, Heidi, et Mary Beth Rosson. 2013. « Using scaffolded examples to teach computational thinking concepts ». P. 95-100 in.
14. Kim, ChanMin, Jiangmei Yuan, Lucas Vasconcelos, et Minyoung Shin. 2018. « Debugging during block-based programming ». *Instructional Science* 46. doi: 10.1007/s11251-018-9453-5.
15. Tashakkori, Abbas, et Charles Teddlie. 2010. *SAGE Handbook of Mixed Methods in Social & Behavioral Research*. 2455 Teller Road, Thousand Oaks California 91320 United States: SAGE Publications, Inc.
16. Biesta, Gert. 2010. « Pragmatism and the philosophical foundations of mixed methods research ».
17. Le Du, Jérémy, Julian Alvarez, et Daniel Schmitt. 2024. « Développer la pensée informatique à travers la conception de jeux vidéo éducatifs Educational Game & Play Design

Project : élaboration de la méthode ». *Distance et Médiation des savoirs* (48). doi: https://doi.org/10.4000/12xon.
18. Campe, Shannon, Jill Denner, Emily Green, et David Torres. 2020. « Pair Programming in Middle School: Variations in Interactions and Behaviors ». *Computer Science Education* 30(1):22-46. doi: 10.1080/08993408.2019.1648119.
19. Mayer, Richard. 2018. « Thirty Years of Research on Online Learning ». *Applied Cognitive Psychology* 33. doi: 10.1002/acp.3482.
20. Schmitt, Daniel, et Olivier Aubert. 2017. « REMIND : une méthode pour comprendre la micro-dynamique de l'expérience des visiteurs de musées ». *RIHM, Revue des Interactions Humaines Médiatisées* 17:43-70.
21. Schindler, Maike, et Achim Lilienthal. 2019. « Students' Creative Process in Mathematics: Insights from Eye-Tracking-Stimulated Recall Interview on Students' Work on Multiple Solution Tasks ». *International Journal of Science and Mathematics Education*. doi: 10.1007/s10763-019-10033-0.
22. Theureau, Jacques. 2006. *Le cours d'action: Méthode développée*. Octares Editions.
23. Romero, Margarida. 2017. « Les compétences pour le XXI e siècle ». in *Usages créatifs du numérique pour l'apprentissage au XXIe siècle*.
24. Le Du, Jérémy, Daniel Schmitt, et Julian Alvarez. Reviewed. « Assessing 21st Century Skills with Stimulated Recall Interviews with Situated Subjective Perspective: The Remind Method ».
25. Román-González, Marcos. 2015. « Computational Thinking Test: Design Guidelines and Content Validation ». P. 2436-44 in *7th International Conference on Education and New Learning Technologies*. Barcelona, Spain.
26. Román-González, Marcos, Juan-Carlos Pérez-González, et Carmen Jiménez-Fernández. 2017. « Which Cognitive Abilities Underlie Computational Thinking? Criterion Validity of the Computational Thinking Test ». *Computers in Human Behavior* 72:678-91. doi: 10.1016/j.chb.2016.08.047.
27. Wiebe, Eric, Jennifer London, Osman Aksit, Bradford W. Mott, Kristy Elizabeth Boyer, et James C. Lester. 2019. « Development of a Lean Computational Thinking Abilities Assessment for Middle Grades Students ». P. 456-61 in *SIGCSE '19: The 50th ACM Technical Symposium on Computer Science Education*. Minneapolis MN USA: ACM.
28. El-Hamamsy, Laila, María Zapata-Cáceres, Estefanía Martín Barroso, Francesco Mondada, Jessica Dehler Zufferey, et Barbara Bruno. 2022. « The Competent Computational Thinking Test: Development and Validation of an Unplugged Computational Thinking Test for Upper Primary School ». *Journal of Educational Computing Research* 60(7):1818-66. doi: 10.1177/07356331221081753.
29. Repenning, Alexander, David C. Webb, Kyu Han Koh, Hilarie Nickerson, Susan B. Miller, Catharine Brand, Ian Her Many Horses, Ashok Basawapatna, Fred Gluck, Ryan Grover, Kris Gutierrez, et Nadia Repenning. 2015. « Scalable Game Design: A Strategy to Bring Systemic Computer Science Education to Schools through Game Design and Simulation Creation ». *ACM Trans. Comput. Educ.* 15(2). doi: 10.1145/2700517.
30. Howland, Kate, et Judith Good. 2015. « Learning to communicate computationally with Flip: A bi-modal programming language for game creation ». *Computers & Education* 80:224-40. doi: 10.1016/j.compedu.2014.08.014.
31. Kim, ChanMin, Lucas Vasconcelos, Brian Belland, Duygu Umutlu, et Cory Gleasman. 2022. « Debugging behaviors of early childhood teacher candidates with or without scaffolding ». *International Journal of Educational Technology in Higher Education* 19. doi: 10.1186/s41239-022-00319-9.

32. Gao, Xuemin, et Khe Hew. 2023. « A Flipped Systematic Debugging Approach to Enhance Elementary Students' Program Debugging Performance and Optimize Cognitive Load ». *Journal of Educational Computing Research* 61:1-32. doi: 10.1177/07356331221133560.
33. Socratous, Chrysanthos, et Andri Ioannou. 2020. « Common Errors, Successful Debugging, and Engagement During Block-Based Programming Using Educational Robotics in Elementary Education ».
34. Bofferding, Laura, Sezai Kocabas, Mahtob Aqazade, et Lizhen Chen. 2022. « The Effect of Play and Worked Examples on First and Third Graders' Creating and Debugging of Programming Algorithms ».
35. Romero, Margarida, Alexandre Lepage, et Benjamin Lille. 2017. « Computational thinking development through creative programming in higher education ». *International Journal of Educational Technology in Higher Education* 14. doi: 10.1186/s41239-017-0080-z.
36. Kafai, Yasmin B., et Quinn Burke. 2015. « Constructionist Gaming: Understanding the Benefits of Making Games for Learning. » *Educational Psychologist* 50(4):313-34. doi: 10.1080/00461520.2015.1124022.

Epistémologie instrumentée en EIAH : retour sur l'outil SEED

Nadine Mandran[1][0000−0002−8660−3827] and Alexis Lebis[2][0000−0003−2104−8671]

[1] Univ. Grenoble Alpes, CNRS, Grenoble INP, LIG, 38000, Grenoble, France
nadine.mandran@univ-grenoble-alpes
[2] IMT Nord Europe, Institut Mines-Télécom, Univ. Lille, Centre for Digital Systems, F-59000, Lille, France alexis.lebis@imt-nord-europe.fr

Résumé Cet article propose l'outil SEED (SEnsibilisation à l'Épistémologie pour les Doctorants), un questionnaire réflexif.Cet outil a pour finalité de sensibiliser les doctorants à l'épistémologie de la recherche grâce à neuf questions. En effet, les objectifs et les concepts de l'épistémologie sont rarement explicités dans les cursus universitaires, et les doctorants adoptent souvent les cadres prédominants de leur discipline. Les réponses aux propositions de SEED élaborent un énoncé et un graphe qui décrivent la posture épistémologique du doctorant. Testé auprès de chercheurs et doctorants, SEED a été jugé pertinent pour susciter une réflexion critique et ouvrir le débat. Cet outil pédagogique semble offrir un support structuré pour guider les doctorants dans leur réflexion épistémologique.

Keywords: épistémologie · doctorant · méthode · sensibilisation · formation

Abstract. This article presents the SEED questionnaire for raising awarness regarding epistemology for doctoral students The purpose of this tool is to raise doctoral students' sensitization of research epistemology through nine questions. Indeed, the objectives and concepts of epistemology are rarely made explicit in academic curricula, and doctoral students often adopt the dominant frameworks of their discipline. The responses to SEED's proposals generate a statement and a graph that describe the epistemological stance of the doctoral student. Tested with researchers and doctoral students, SEED was found to be relevant for stimulating critical reflection. This pedagogical tool seems to offer a structured support to guide doctoral students in their epistemological reflection.

Keywords: epistemology · PHD student · sensitization · training · method

1 Introduction

Les courants épistémologiques trouvent leur fondement dans des questionnements philosophiques [4]. Ces réflexions définissent des cadres pour orienter les pratiques de recherche, en établissant des principes tels que la validité ou la valeur des connaissances scientifiques. Pour le chercheur, ce postulat implique d'identifier comment les connaissances sont constituées et quelles sont leur valeur et leur validité [3]. Cette réflexion est donc une composante essentielle de la formation du chercheur [12]. Pourtant, la diversité des courants épistémologiques est rarement abordée de manière approfondie et, en pratique, les doctorants adoptent, sans en être pleinement conscients, le cadre épistémologique prédominant de leur discipline. Toutefois, c'est cette réflexion qui permet au chercheur de situer ses travaux dans un cadre, de faciliter le dialogue avec d'autres chercheurs issus de postures épistémologiques différentes, de se prémunir de biais de jugement et de défendre la validité de leur contribution. [19].

Malgré la complexité de ces concepts [1], il apparaît essentiel d'initier les doctorants à ces réflexions, car elles les aideront à justifier leur travail. Pour les accompagner dans cette démarche, nous proposons le questionnaire SEED. Cet article présente les difficultés auxquelles se confrontent l'enseignement de l'épistémologie, le questionnaire SEED et les premiers résultats de son utilisation.

2 Etat de l'art

L'enseignement de l'épistémologie est rarement proposé comme une discipline à part entière dans les cursus de masters, notamment en dehors des domaines philosophiques ou des disciplines où les courants épistémologiques sont intégrés par les communautés scientifiques [28] (e.g. la réfutabilité selon T. Kuhn [11] ou les écoles de pensée en sociologie [13]). Dans l'enseignement, l'accent est principalement mis sur l'enseignement de méthodes de production des données (qualitatif/quantitatif), reléguant l'épistémologie à des références implicites, sous-jacentes [2],[12]. Ce déficit de connaissances épistémologiques limite les doctorants dans leur compréhension et leur pratique du métier de chercheur. Mal préparés à répondre à des critiques ou à interroger les présupposés des méthodes qu'ils emploient, ils manquent souvent de recul critique et de réflexion approfondie sur la nature des connaissances qu'ils produisent [6]. En outre, ce manque entrave les interactions pluridisciplinaires. En effet, les divergences méthodologiques entre disciplines, au lieu de générer des complémentarités, risquent de créer des oppositions limitant la collaboration [10].

La littérature identifie plusieurs freins à l'enseignement de l'épistémologie.

- Un enseignement des méthodes au détriment de l'épistémologie : « Dans de nombreuses disciplines, il y a une confusion fréquente entre méthode et épistémologie : les outils méthodologiques sont enseignés comme des pratiques normatives, sans que les étudiants soient invités à en interroger les fondements épistémologiques. » [5]. L'épistémologie se transmet implicitement via les cours sur les méthodes ou les cadres théoriques, sans nécessiter

un enseignement spécifique. Or, pour nous, c'est réducteur, car la posture épistémologique justifie les méthodes de conduite de la recherche et pas l'inverse.
- Peu de temps dédié à cette discipline : les masters de recherche disposent de peu de crédits alloués à ces enseignements théoriques, privilégiant la spécialisation disciplinaire et les compétences techniques. Les savoirs transversaux comme l'épistémologie peuvent être perçus comme secondaires par rapport à l'acquisition de compétences disciplinaires.
- Nature abstraite et langage théorique spécialisé : l'épistémologie est une réflexion sur la science qui traite de concepts fondamentaux sur la construction des savoirs « L'épistémologie est une discipline exigeante qui oblige à prendre du recul sur les savoirs et à questionner les évidences. Cette double démarche, à la fois critique et réflexive, la rend particulièrement complexe à transmettre ».[1]
- Ancrage transmissif : L'enseignement de l'épistémologie repose souvent sur l'explication des courants philosophiques et épistémologiques. Des cas concrets viennent parfois illustrer ces grands courants. Les outils réflexifs pour accompagner cette démarche restent rares [28]. Ortiz-Revilla, J. et al (2020) (p.873) [27] proposent une grille listant les caractéristiques descriptives de l'épistémologie mais sans toutefois stimuler un questionnement critique (e.g. par des questions).

Face à ces constats, comment pouvons-nous sensibiliser les doctorants à cette discipline complexe mais fondatrice du métier de chercheur ?

3 Contribution

Notre objectif est de sensibiliser les doctorants à l'épistémologie pour renforcer leur esprit critique et leur ouverture disciplinaire via le questionnaire SEED (SEnsibilisation à l'Epistémologie pour les Doctorants) et un scénario pédagogique.
SEED se focalise sur 4 courants épistémologiques. Le post-positivisme (P) tout en admettant une réalité objective, reconnaît la partialité et l'incertitude de la connaissance, valorise la falsifiabilité des théories, adopte une approche critique et contextuelle, et prône une pluralité de méthodes. [9], [14]. Le réalisme critique (R) reconnaît qu'il existe une réalité objective indépendante des perceptions humaines, constituée de structures et mécanismes sous-jacents. Cependant, il admet que notre accès à cette réalité est toujours partiel, car il est médiatisé par nos cadres conceptuels et contextes sociaux. Cette posture cherche à dévoiler les causes profondes des phénomènes observables. Elle adopte une démarche scientifique critique et réflexive [7],[8], [25], [29]. Le constructivisme pragmatique (C) valorise les processus par lesquels les significations et les vérités sont co-construites, tout en se demandant si ces vérités fonctionnent dans la pratique ou aident à produire des résultats pertinents dans un contexte donné. Plutôt que de chercher des vérités absolues, cette approche se concentre sur la manière

dont ces vérités émergent dans des contextes spécifiques. Elle met un accent particulier sur l'utilité de ces vérités, en évaluant si elles produisent des résultats efficaces dans des situations concrètes. [15], [16], [22], [23]. L'interprétativisme (I) adopte une posture subjectiviste par rapport à la réalité. Il considère que La réalité sociale n'est pas une entité objective existant indépendamment des individus. Elle est construite à travers les significations, les interprétations et les expériences des acteurs sociaux. Ainsi, la réalité est multiple et contextuelle, dépendante des perceptions et des interactions humaines plutôt que d'une vérité universelle [26], [17], [18], [20], [21]. Ces paradigmes ont été choisis pour leur diversité dans l'approche de la complexité du réel, leur positionnement sur des faits "objectifs" ou socialement construits, et leur représentativité dans la communauté EIAH.
SEED comporte 9 questions fermées et 29 propositions de réponses (3 ou 4 par question) basées sur ces quatre paradigmes épistémologiques. Ci-après, nous présentons les couples questions/propositions :

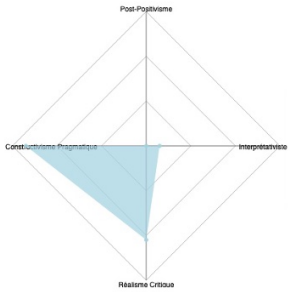
1. **Concernant votre travail, comment définiriez-vous la nature de la réalité ?**
   - La réalité existe indépendamment de l'humain (P)
   - La réalité est en partie construite socialement et subjectivement interprétée (R) et (C)
   - Des faits socialement construits non dirigés par des lois naturelles existent (I)
2. **Concernant votre travail, quelle est votre opinion sur la possibilité de connaître la réalité ?**
   - Le réel est connu avec une éventuelle défaillance des instruments de mesure (P)
   - Le domaine réel n'est pas observable, des événements sont observables (R)
   - L'expérience humaine est observable dans le système auquel elle appartient (C)
   - Des faits sont produits comme une partie des interactions sociales entre des chercheurs et des individus (I)
3. **Concernant votre travail, quel est l'objectif de la connaissance scientifique produite ?**
   - L'objectif de la connaissance scientifique est d'identifier des régularités, des invariants. (P)
   - L'objectif de la connaissance scientifique est d'identifier des mécanismes générateurs, des modèles. (R)
   - L'objectif de la connaissance scientifique est de construire des modèles intelligibles de l'expérience humaine offrant des repères, des outils adaptés. (C)
   - L'objectif de la connaissance scientifique est de comprendre comment les individus donnent du sens aux expériences sociales qu'ils vivent. (I)
4. **Concernant votre travail, comment percevez-vous le rôle des valeurs personnelles et des croyances dans la formation des connaissances ?**

   - Les valeurs personnelles doivent être mises de côté pour parvenir à une compréhension objective de la réalité. (P)
   - Les valeurs personnelles influencent inévitablement nos perspectives et nos interprétations. (RC)
   - Les valeurs personnelles sont nécessaires pour construire et interpréter les connaissances. (I)
5. **Concernant votre travail, quelle est votre attitude envers les biais et la subjectivité dans la connaissance scientifique ?**
   - La subjectivité doit être minimisée pour parvenir à une compréhension objective de la réalité. (P)
   - Les biais sont inévitables et il est important de les reconnaître et de les gérer. (RC)
   - La subjectivité fait partie de la construction de la connaissance. (I)
6. **Concernant votre travail, quelle est votre attitude envers le rôle des contextes sociaux et historiques dans la production de la connaissance ?**
   - Les contextes sociaux et historiques sont secondaires pour la construction de la connaissance. (P)
   - Les contextes sociaux et historiques sont importants pour la construction de la connaissance. (RC)
   - Les contextes sociaux et historiques sont fondamentaux pour la construction de la connaissance. (I)
7. **Concernant votre travail, quel matériau empirique est mobilisé pour construire la connaissance ?**
   - Pour construire et valider des connaissances scientifiques, des données quantitatives sont élaborées à partir de plans expérimentaux contrôlés. (P)
   - Pour construire et évaluer des connaissances scientifiques, l'observation d'évènements dans différents contextes est nécessaire (RC)
   - Pour construire et évaluer la connaissance scientifique, des narrations auprès d'individus sont recueillies (I)
8. **Concernant votre travail, sur quoi repose la fiabilité du processus de production et d'analyse des données ?**
   - La fiabilité de production et d'analyse des données repose sur la capacité à pouvoir reproduire les résultats expérimentaux. La réplication et la fiabilité des mesures reposent sur la description du protocole et des données. (P)
   - La fiabilité de production et d'analyse des données repose sur la description du cheminement cognitif pour construire la connaissance et sa traçabilité. (RC)
   - La fiabilité de production et d'analyse des données repose sur le recueil de nombreuses narrations et la manière dont les interprétations des narrations sont vérifiées. (I)
9. **Concernant votre travail, comment la connaissance est-elle mise à l'épreuve ?**

– Des tests statistiques sont réalisés pour valider ou invalider des hypothèses, permettant de tester la connaissance. (P)
– Pour éprouver la connaissance, le modèle et ses effets doivent être testés dans des contextes différents. (RC)
– Il n'y a pas besoin de mettre à l'épreuve la connaissance une fois produite. (I)

Les propositions de réponses apparaissent dans un ordre aléatoire pour éviter les biais systématiques, et les paradigmes restent cachés pour favoriser une réflexion sur chaque proposition plutôt qu'un alignement automatique.

L'outil construit progressivement une connaissance de la posture du répondant, restituée en trois formes : 1) une description concise de la posture épistémologique dominante, à confirmer ou non par le répondant ; 2) un diagramme de Kiviat [3] illustrant graphiquement le positionnement sur les quatre postures, basé sur une agrégation des réponses pondérées ; et 3) un texte reflétant les choix et résumant la posture du répondant. , Ce texte formulé à la première personne renforce l'aspect réflexif du questionnaire Figure 1.



Je pense que la réalité existe indépendamment de l'humain, que l'expérience humaine est observable dans le système auquel elle appartient, que l'objectif de la connaissance scientifique est d'identifier des mécanismes générateurs, que la connaissance scientifique est produite à travers un processus de co-construction sociale, où les interactions entre les chercheurs et les individus concourent à cette construction. Il me semble que les valeurs personnelles sont nécessaires pour construire et interpréter les connaissances, que les biais sont inévitables et il est important de les reconnaître et de les gérer, que les contextes sociaux et historiques sont fondamentaux pour la construction de la connaissance. Il me semble que pour construire et valider des connaissances scientifiques, des données quantitatives sont élaborées à partir de plans expérimentaux contrôlés. Pour finir, il m'apparaît que la fiabilité de production et d'analyse des données repose sur le recueil de nombreuses narrations et la manière dont les interprétations des narrations sont vérifiées.

Figure 1. Exemple de texte et de graphe proposés à la fin du questionnaire

L'outil SEED est disponible en ligne [24]. Il est utilisable en local avec un navigateur et ne requiert aucune installation. SEED est archivé sur Software Heritage.

4 Conception et mise à l'épreuve

SEED a été développé en plusieurs itérations. La première version, inspirée de M.J. Avenier et C. Thomas (2015) [3], comprenait quatre volets : 1) Hypothèses ontologiques et épistémiques, 2) But, statut et forme de la connaissance, 3) Validité : matériau empirique et fiabilité, et 4) Valeur : qualité des construits et

[3] Créé en 1877 par le statisticien Georg von Mayr.

contribution. Cependant, ces concepts, enseignés dans des formations doctorales, restaient abstraits. Un atelier (Leysin, 2022) et un séminaire avec 5 chercheurs ont révélé des biais et une rigidité dans les propositions. Nous avons alors utilisé des Large Language Models[31], via ChatGPT, pour reformuler et simplifier les questions selon quatre courants épistémologiques. Les requêtes, répétées cinq fois dans des contextes neutres, ont permis d'élaborer une version affinée de SEED après vérification et ajustement. Ce processus, mené avec 2-3 chercheurs à chaque étape, a été répété trois fois. Ensuite, nous avons évalué SEED auprès de quatre chercheurs de disciplines variées (Mécanique des fluides, Recherche opérationnelle, Sciences de l'éducation). Après une introduction à l'épistémologie et aux objectifs de l'étude, ils répondaient au questionnaire de manière autonome. Ensuite, un entretien explorait 1) leur avis sur la position épistémologique proposée, 2) la clarté des questions, et 3) leurs suggestions d'amélioration. SEED a également été utilisé lors d'une formation auprès de 12 doctorants de disciplines différentes (La Rochelle - octobre 2024). Nous avons recueilli leur avis sur la clarté des questions, avec une échelle allant de 0 à 4 (0 : proposition pas du tout claire à 4 : proposition tout à fait claire).

5 Résultats

Lors des entretiens avec les chercheurs, il s'est avéré que SEED est un outil "très facile et qui permet la réflexion" (MG)[4] et "on retrouve bien les postures dans les phrases du questionnaire" (BL). Ce travail réflexif et guidé permet "de savoir ce que l'on veut faire, c'est une boussole pour trouver la méthode de recherche" (BL). Le texte proposé pour résumer la posture est perçu comme "bluffant" (MG), "résume très clairement la posture. Je me suis reconnu" (BL) et "Semble bien cerner ma conception des choses" (TL).

L'ensemble des répondants auraient souhaité pouvoir choisir entre plusieurs propositions "C'est difficile de choisir une seule réponse" (MG). Pour deux personnes, il faudrait offrir un moyen de hiérarchiser les réponses (MG, BL) et de supprimer certains items non pertinents pour eux (MG), ce qui permettrait éventuellement de rajouter plus de nuance "des fois j'avais envie de mettre plusieurs réponse ou d'ajouter de la nuance" (FL).

Les participants ont insisté sur l'importance d'adapter la réflexion au travail personnel de recherche (MG, BL). Ainsi, nous avons ajouté "Concernant votre travail" au début de chaque question pour recentrer l'utilisateur sur son propre contexte.

SEED apporte une aide sur le positionnement épistémologique en cela qu'un des participants a déclaré qu'il n'avait "Aucune" (TL) connaissance par rapport à sa posture épistémologique et que "Oui !" (TL), à l'issue de l'utilisation de l'outil il avait appris sur cette thématique. Ce chercheur propose d'ailleurs d'utiliser cet outil pour aider les doctorants afin de les "inviter à se poser des questions fondamentales sur la démarche scientifique, même si [il] ne pense pas que cela impactera leur manière de conduire leur thèse à court terme" (TL).

[4] Le code entre parenthèses indique l'auteur du propos

Avec les réponses des doctorants pour mesurer la clarté des propositions, nous obtenons des résultats satisfaisants : la moyenne pour l'ensemble des propositions est de 3,6/4, le minimum de 3,1 et le maximum de 4. Les propositions les moins bien perçues sont : "Des faits sont produits comme une partie des interactions sociales entre des chercheurs et des individus (I)" : score 3,1/4 "L'objectif de la connaissance scientifique est de construire des modèles intelligibles de l'expérience humaine offrant des repères, des outils adaptés." (C) : score 3,3/4 "L'objectif de la connaissance scientifique est de comprendre comment les individus donnent du sens aux expériences sociales qu'ils vivent. (I)" : score 3,3/4 "La fiabilité de production et d'analyse des données repose sur la description du cheminement cognitif pour construire la connaissance et sa traçabilité." (RC) : score 3,3/4.

Ces résultats nous ont conduit à formuler de nouvelles propositions de réponse afin de limiter les ambiguïtés et de faciliter la compréhension de ces concepts complexes . Nous devons continuer à les évaluer avec d'autres indicateurs et avec plus de personnes.

6 Conclusion

Face aux obstacles rencontrés par les doctorants pour se positionner sur un courant épistémologique, nous avons élaboré le questionnaire SEED. Ce dernier semble aider les doctorants à clarifier leurs choix épistémologiques en proposant un positionnement et sa description. Les chercheurs interrogés ont jugé SEED pertinent pour réfléchir à la constitution des connaissances scientifiques. Les premiers retours des doctorants montrent que les propositions du questionnaire ont été comprises mais des évaluations complémentaires sont nécessaires. En terme de perspectives, il serait enrichissant de tester l'impact du questionnaire dans le dispositif d'encadrement et sa complémentarité avec des outils permettant d'aligner les attentes entre encadrants et doctorant [30]. Par exemple, la posture proposée par SEED pourrait guider les doctorants dans le choix de méthodes de conduite de recherche adaptées à leur contexte. Ainsi, SEED pourrait devenir une ressource pour répondre aux enjeux méthodologiques des parcours doctoraux.

Remerciements. Cette contribution scientifique a été réalisée dans le cadre du projet de recherche TALE4GDA, financé par l'ANR (ANR-23-CE38-0001). Nous remercions les chercheurs et doctorants qui nous ont aidé à concevoir et évaluer SEED.

Références

1. Astolfi, J.P. : L'école pour apprendre : L'élève face aux savoirs. Esf (2010)
2. Avenier, M.J., Thomas, C. : Mixer quali et quanti pour quoi faire ? Méthodologie sans épistémologie n'est que ruine de réflexion (Dec 2011), `https://halshs.archives-ouvertes.fr/halshs-00644303/document`

3. Avenier, M.J., Thomas, C. : Finding one's way around various methodological guidelines for doing rigorous case studies : A comparison of four epistemological frameworks. Systèmes d'information & management **20**(1), 61–98 (2015), http://www.cairn.info/resume.php?ID_ARTICLE=SIM_151_0061
4. Avrum, S., Martinichi, A. : Epistémology (Oct 2024), https://www.britannica.com/topic/epistemology
5. Bachelard, G. : La Formation de l'esprit scientifique. Librairie philosophique, Paris, 5e édition edn. (1967)
6. Bangali, M. : Les acquis de la formation doctorale : perceptions des compétences développées. Canadian Journal of Higher Education / Revue canadienne d&#x27 ;enseignement supérieur **51**(1), 15–27 (2021). https://doi.org/https://doi.org/10.47678/cjhe.vi0.188885, publisher : Canadian Society for the Study of Higher Education
7. Bhaskar, R. : Philosophy and scientific realism. In : Critical realism, pp. 16–47. Routledge (2013)
8. Bhaskar, R. : A realist theory of science. Routledge (2013)
9. Boisot, M., McKelvey, B. : Integrating modernist and postmodernist perspectives on organizations : A complexity science bridge. Academy of management review **35**(3), 415–433 (2010), publisher : Academy of Management Briarcliff Manor, NY
10. Boix-Mansila, V. : Interdisciplinary work at the frontier : An empirical examination of expert interdisciplinary epistemologies. Issues in integrative studies (24), 1–31 (2006)
11. Chalmers, A. : What is this thing called science ? McGraw-Hill Education (UK) (2013)
12. Dumez, H., Grandazzi, A., Senn, J. : Chapitre 7. Faut-il faire de l'épistémologie ? Et si oui, comment ? In : L'expérience de la thèse en management, pp. 138–148. Gestion en liberté, EMS Éditions, Caen (2023). https://doi.org/10.3917/ems.gaill.2023.01.0138, https://shs.cairn.info/l-experience-de-la-these-en-management--9782376874881-page-138?lang=fr
13. Freitag, M. : Pour un dépassement de l'opposition entre "holisme" et "individualisme" en sociologie. Revue européenne des sciences sociales **32**(99), 169–219 (1994), http://www.jstor.org/stable/40370930, publisher : Librairie Droz
14. Gephart Jr, R.P. : Qualitative research and the Academy of Management Journal (2004), issue : 4 Pages : 454–462 Publication Title : Academy of management journal Volume : 47
15. von Glasersfeld, E. : An introduction to radical constructivism. The invented reality **1740**, 28 (1984), publisher : New York, NY : Norton
16. von Glasersfeld, E. : The radical constructivist view of science. Foundations of science **6**(1-3), 31–43 (2001), http://link.springer.com/article/10.1023/A:1011345023932
17. Guba, E.G., Lincoln, Y.S. : Fourth generation evaluation. Sage (1989)
18. Guba, E.G., Lincoln, Y.S. : Paradigmatic controversies, contradictions, and emerging confluences. Sage Publications Ltd (2005)
19. Katie Moon, Blackman, D. : A guide to ontology, epistemology, and philosophical perspectives for interdisciplinary researchers (May 2017), https://i2insights.org/2017/05/02/philosophy-for-interdisciplinarity/
20. Klein, H.K., Myers, M.D. : A set of principles for conducting and evaluating interpretive field studies in information systems. MIS quarterly pp. 67–93 (1999), publisher : JSTOR

21. Klein, H.K., Rowe, F. : Marshaling the professional experience of doctoral students : A contribution to the practical relevance debate. MIS Quarterly pp. 675–686 (2008), publisher : JSTOR
22. Le Moigne, J.L. : Le constructivisme, tome 1 : Les enracinements. Paris : L'Harmattan (2001)
23. Le Moigne, J.L. : Repères historiques pour l'entendement des épistémologies constructivistes. Que sais-je ? **3**(2969), 41–69 (2007), publisher : Presses Universitaires de France
24. Lebis, A. : Epistemological self-assessment tool (May 2024), https://archive.softwareheritage.org/swh:1:dir:5c38fab90e1352a7dbba418816fd5b2bf7355262
25. Mingers, J., Mutch, A., Willcocks, L. : Critical realism in information systems research. MIS quarterly **37**(3), 795–802 (2013), publisher : JSTOR
26. Orlikowski, W.J., Baroudi, J.J. : Studying information technology in organizations : Research approaches and assumptions. Information systems research **2**(1), 1–28 (1991), publisher : INFORMS
27. Ortiz-Revilla, J., Adúriz-Bravo, A., Greca, I.M. : A Framework for Epistemological Discussion on Integrated STEM Education. Science & Education **29**(4), 857–880 (Aug 2020). https://doi.org/10.1007/s11191-020-00131-9, https://doi.org/10.1007/s11191-020-00131-9
28. Ryder, J., Hind, A., Leach, J. : Teaching about the Epistemology of Science in School Science Classrooms : Case Studies of Teachers' Experiences. In : Boersma, K., Goedhart, M., de Jong, O., Eijkelhof, H. (eds.) Research and the Quality of Science Education, pp. 283–293. Springer Netherlands, Dordrecht (2005). https://doi.org/10.1007/1-4020-3673-6_23, https://doi.org/10.1007/1-4020-3673-6_23
29. Smith, M.L. : Overcoming theory-practice inconsistencies : Critical realism and information systems research. Information and organization **16**(3), 191–211 (2006), publisher : Elsevier
30. Stappenbelt, B., Basu, A. : Student-supervisor-university expectation alignment in the undergraduate engineering thesis. Journal of Technology and Science Education **9**(2), 199–216 (2019). https://doi.org/10.3926/jotse.482, https://www.jotse.org/index.php/jotse/article/view/482
31. Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M.A., Lacroix, T., Rozière, B., Goyal, N., Hambro, E., Azhar, F., et al. : Llama : Open and efficient foundation language models. arXiv preprint arXiv :2302.13971 (2023)

De l'itératif à la navigation accompagnée : une évolution des méthodes de conduite de la recherche

Nadine Mandran[1][0000−0002−8660−3827], Alexis Lebis[2][0000−0003−2104−8671], and Mathieu Vermeulen[2][0000−0003−3646−1741]

[1] Univ. Grenoble Alpes, CNRS, Grenoble INP, LIG, 38000, Grenoble, France
nadine.mandran@univ-grenoble-alpes
[2] IMT Nord Europe, Institut Mines-Télécom, Univ. Lille, Centre for Digital Systems, F-59000, Lile, France alexis.lebis@imt-nord-europe.fr

Résumé Pour aider les doctorants à naviguer dans des méthodes de recherche itératives souvent rigides, cet article propose la méta-méthode T5S (THEDRE in 5 Spaces). Repensant les processus linéaires, T5S permet de passer librement entre cinq espaces interconnectés : argumentation, questionnement, création, mise à l'épreuve, et échanges. Inspirée de THEDRE et développée à partir d'ateliers avec des chercheurs, T5S a été évaluée auprès de 32 doctorants.

Mots-clés : méthode de recherche · recherche doctorale · flexibilité · traçabilité · collaboration · iteration

Abstract. To help doctoral students navigate often rigid iterative research methods, this article proposes the T5S meta-method (THEDRE in 5 Spaces). Rethinking linear processes, T5S allows free movement between five interconnected spaces: argumentation, questioning, creation, testing, and exchange. Inspired by THEDRE and developed through workshops with researchers, T5S was evaluated with 32 doctoral students.

Keywords: research method · doctoral research · flexibility· collaboration · iteration

1 Introduction

Dans les recherches de type recherche-développement ou recherche-action, le travail de recherche reste souvent opaque pour les doctorants. Leur formation, centrée sur des aspects théoriques, manque de pragmatisme, ce qui limite leur compréhension des exigences du métier de chercheur. Ils peinent à prendre du recul sur leurs contributions et à s'orienter [3]. Bien que l'itération soit reconnue comme

[3] https ://www.sciencedirect.com/science/article/abs/pii/S1775878524006933

essentielle, elle reste peu formalisée : quand et pourquoi itérer ? Un processus itératif en conception ou recherche vise à affiner un modèle ou outil à chaque cycle [16]. C'est une caractéristique clé des approches comme le *Human-Centered Design* (HCD) [1] et le *Design Thinking* (DT) [4], qui prônent des cycles permettant des révisions basées sur les retours d'utilisateurs. Des méthodes comme la Recherche en Informatique Centrée Humain (RICH) [13], le *Design-Based Research* (DBR) [20], la Recherche Orientée par la Conception (ROC) [18], et THEDRE [13] suivent cette vision itérative. Cependant, elles restent limitées par un enchaînement rigide d'étapes. Par exemple, THEDRE s'appuie sur le cycle de Deming [9], tandis que DBR propose une itération sans outil pour en gérer les cycles. Ainsi, pour former les doctorants à la recherche, il semble important d'éclaircir cette question du processus itératif de la recherche et d'identifier à quel moment la prise de recul sur le travail est nécessaire. Cet article aborde les questions suivantes : Comment formaliser la succession des étapes dans un processus de recherche en lien avec le terrain ? Comment formaliser la prise de recul et les échanges avec les pairs ?
Après avoir présenté les principes itératifs de différentes méthodes, nous proposons la méta-méthode T5S (THEDRE in 5 Spaces) caractérisée par sa structure non linéaire en 5 espaces non ordonnés.

T5S se distingue par l'intégration d'un espace dédié aux échanges avec les encadrants et la communauté scientifique. Cette méthode conçue avec un groupe de chercheurs (septembre 2023 - deux focus-groups) a fait l'objet d'une première mise en situation lors de formations dans des écoles doctorales (octobre et novembre 2024 - 33 étudiants).

2 État de l'art

Nous analysons trois méthodes de conception itérative : le *Human-Centered Design* (HCD) [5], [1], le *Design Thinking* (DT) et le modèle du Double Diamant (DD) [10] et trois méthodes de conduite de la recherche : le *Design Based Research*, la recherche orientée par la conception (ROC) et la méthode THEDRE. Nous avons fait ce choix car ces méthodes se préoccupent d'intégrer l'humain au centre de la recherche. Trois questions guident l'analyse : Quelles sont les étapes proposées par ces méthodes ? Comment les itérations/transitions entre ces étapes sont-elles gérées ? Des outils sont-ils proposés pour suivre ces étapes ?

2.1 Trois méthodes de conception centrée utilisateurs

Le HCD et le DT intègrent l'utilisateur final tout au long du processus. Elles reposent sur un processus itératif pour s'assurer que les solutions développées sont adaptées aux utilisateurs finaux. Le HCD est décrit en trois étapes : analyse des besoins, co-conception et évaluation [1]. Ces trois étapes s'enchaînent mais les facteurs déclenchant les changements d'étapes et les itérations ne sont pas précisés. Le *Design Thinking* est structuré en 6 étapes [4, 10] : 1)Empathie

(*Empathize*), 2) Définition (*Define*), 3) Idéation (*Ideate*), 4) Prototypage (*Prototype*), 5) Test, 6) Implémentation(*Implement*). La figure 1 du *Nielsen Norman Group* présente ces étapes.

T ·echerche

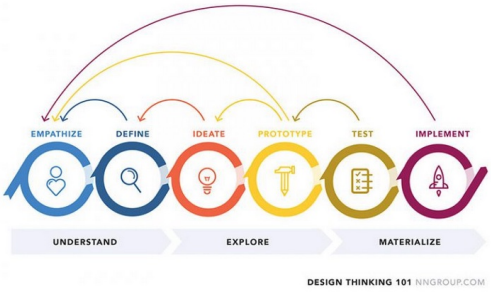


Figure 1. Design Thinking. Source https ://www.usabilis.com/quest-ce-que-le-design-thinking/

Malgré l'apparente flexibilité, les étapes sont organisées de manière séquentielle et les allers-retours contraints. Le modèle DD développé par le *Design Council* en 2005 [7] tente de lever ces limites. Il est représenté sous la forme de deux "diamants" connectés (figure 2).

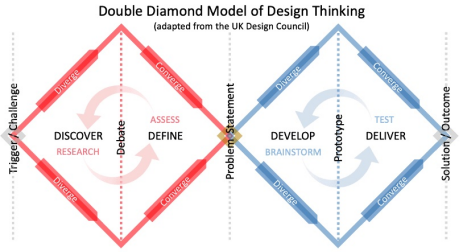


Figure 2. Modèle du Double Diamant du Design Thinking

Le premier, utilisé pour identifier le problème, est composé de deux étapes : *Discover* et *Define*. Le second, pour développer la solution, est composé de deux étapes : *Develop* et *Test*. Chaque diamant est composé de périodes de pensées divergentes et de pensées convergentes [8]. Le DD offre la possibilité de naviguer entre des étapes d'exploration et de solutions. Mis à l'épreuve par Lefebvre, P. (2018) [11], le DD est jugé difficile à suivre :

Les démarches de design observées sur nos cas réels s'écartent du modèle du DD, sur les deux aspects cruciaux de ce modèle : la séquence problème-diagnostic-solution, l'alternance divergence/convergence.

Pour Lefebvre, P. (2018), les séquences éloignées de la réalité ne sont pas assez opérationnelles.
Ces méthodes (HCD,DT,DD) qualifiées d'itératives gardent une structure sous-jacente linéaire pouvant limiter la flexibilité du processus [21]. Aussi, Ces méthodes de conception de dispositifs négligent des étapes de la recherche, comme la phase de problématisation.

2.2 Trois méthodes de conduite de la recherche

Pour conduire une recherche en concevant un dispositif et en intégrant l'ensemble des acteurs éducarionnels, le DBR [20] [6] propose 9 principes. Ils concernent l'organisation de la recherche, l'implication des acteurs et la prise en compte du terrain, la documentation et la généralisation [14]. Un principe concerne l'itération : « La contribution scientifique et les outils associés doivent être revus régulièrement »

Pour opérationnaliser les principes du DBR, McKenney S. et Reeves T. (2018) [15] décrivent un processus linéaire avec 6 étapes avec des *Sub-cycles* à l'intérieur de ces étapes ou des *Multiple sub-cycles* pour plusieurs étapes (figure 3).

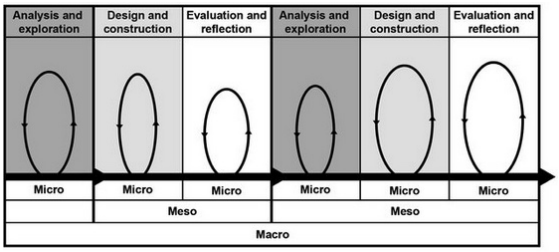


Figure 3. Modèle itératif du DBR de Mc Kenney et Reeves 2018 [15] - p 84

Ce modèle montre la nécessité d'itérer à l'intérieur des étapes et entre chacune d'elles.

La ROC, modélisation du DBR, intercale deux cycles [19] : la conception et l'évaluation comme temps de la recherche (figure 4).

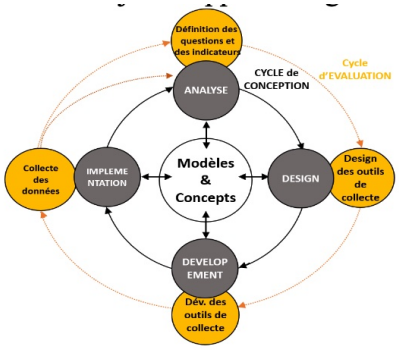


Figure 4. Recherche Orientée par la Conception (ROC) [19] - p 81

Malgré ses allers-retours possibles, cette approche reste contrainte par des étapes fixes. En se centrant sur la conception et l'évaluation, elle néglige des phases clés comme l'exploration, la conceptualisation et la théorisation, ce qui peut rigidifier la distinction entre recherche et conception et limiter l'intégration de nouvelles connaissances.

THEDRE, issue de suivis méthodologiques de thèses dans diverses disciplines, est une méta-méthode itérative en 4 étapes (Planification, Expérimentation, Contrôle, Construction/Décision) pour structurer la recherche. Adaptable à des contextes variés, elle fournit des guides pour accompagner les doctorants. Une version spécifique outille les principes du DBR [14]. Avec le langage de THEDRE, les doctorants peuvent créer leur propre processus de recherche . THEDRE inclut des indicateurs d'objectifs pour gérer les cycles, mais leur choix restent complexes.

Si les méthodes étudiées mettent en avant l'itération, le processus sous-jacent reste une succession linéaire d'étapes. Cela révèle une tension entre la nécessité de l'itération et la réalité de la mise en œuvre de ces méthodes. THEDRE propose un guide pour définir des indicateurs d'objectifs, mais leur choix reste complexe. Les moments d'échanges et de recul, bien que suggérés, ne sont pas non plus formalisés.

3 T5S : Une évolution de THEDRE en 5 espaces

La méta-méthode T5S structure le processus autour de cinq espaces essentiels, indépendants mais interconnectés, favorisant une certaine adaptabilité. T5S se positionne comme une démarche descriptive et non prescriptive.

3.1 Description des 5 espaces

L'espace de l'argumentaire correspond à l'élaboration de la problématique. C'est ici que les questions de recherche se construisent à partir d'un constat

global sur un problème donné. L'objectif est de démontrer l'importance de ce problème et de sa complexité à le résoudre. Il s'agit ensuite de synthétiser les travaux sur le sujet, d'identifier des manques, de définir des questions de recherche non résolues et de se positionner. Cet espace constitue la justification de l'ensemble du travail de recherche.

L'espace du questionnement correspond à l'activité de veille, c'est-à-dire une « activité continue [...] visant à une surveillance active de l'environnement, pour en anticiper les évolutions » [2]. Cette veille nécessite l'étude approfondie de la littérature scientifique et une veille sociétale pour mesurer l'importance du travail et son impact sur la société, voire de consulter des bases de brevets ou de dépôts d'application. Cela permet d'identifier des méthodes reconnues pour construire ou évaluer la contribution. Une revue systématique de la littérature et la documentation de cette dernière garantissent la transparence sur le choix des références utilisées [17, 12].

L'espace de la création est dédié à la conception de la contribution pour répondre aux questions de recherche. L'idée latente de la contribution peut émerger de manière spontanée. Elle peut également résulter d'une construction progressive, à travers une succession d'étapes de réflexion, d'expérimentations, de doutes et d'erreurs. Ce processus peut être individuel ou collectif, les discussions et les critiques des pairs jouant un rôle clé dans l'enrichissement de l'idée. Cet espace implique une complexité cognitive qui mêle connaissances initiales, intuition, esprit critique et apports externes [3] : il est de fait essentiel pour une méthode de conduite de la recherche à destination des doctorants – chercheur en devenir et par conséquent créateur et innovateur.

L'espace de la mise à l'épreuve correspond à la confrontation du problème au terrain et à des tests pour évaluer la contribution. Il peut s'agir d'observation dans un contexte donné pour comprendre et étudier en profondeur un problème. Il peut s'agir aussi de tests, in situ ou en laboratoire, pour éprouver ou évaluer la contribution. Il n'est pas besoin de détailler plus cet espace car il est largement décrit dans une littérature abondante.

L'espace de l'échange correspond aux échanges entre le doctorant et les différents acteurs de la recherche, comme la collaboration entre le doctorant et son encadrement, ou les discussions avec les pairs. C'est également là où s'opère la prise de recul du doctorant. Cet espace, incluant les encadrants, favorise une explicitation des idées et une réflexion critique sur le travail en cours. Expliciter cet espace est important pour permettre au doctorant de se rendre compte de la nécessité de la prise de recul sur son travail, de la clarification de ses propos afin d'échanger plus facilement avec ses encadrants et identifier dans quelle voie le travail continuera. Pour guider ce travail, T5S propose des questions réflexives à destination du doctorant, comme : Ma contribution est-elle suffisamment aboutie pour être mise à l'épreuve ? ; La problématique est-elle suffisamment argumentée ? ; Dois-je à nouveau consulter la littérature ? En l'absence d'un tel espace dans un travail doctoral, le risque est de voir le doctorant s'enfermer dans un sujet, d'explorer des éléments hors du sujet, de se disperser ou de ne pas être en

mesure d'argumenter et de prendre du recul. La figure 5 présente l'organisation des 5 espaces.

T5S se distingue des approches linéaires ou itératives par la porosité entre ses cinq espaces. Cette flexibilité permet une navigation adaptative entre les espaces, notamment face aux imprévus, sans ordre prédéfini ni séquentialité (figure 5). L'accent est mis sur les transitions entre espaces, qui reflètent la réalité dynamique du travail de recherche.

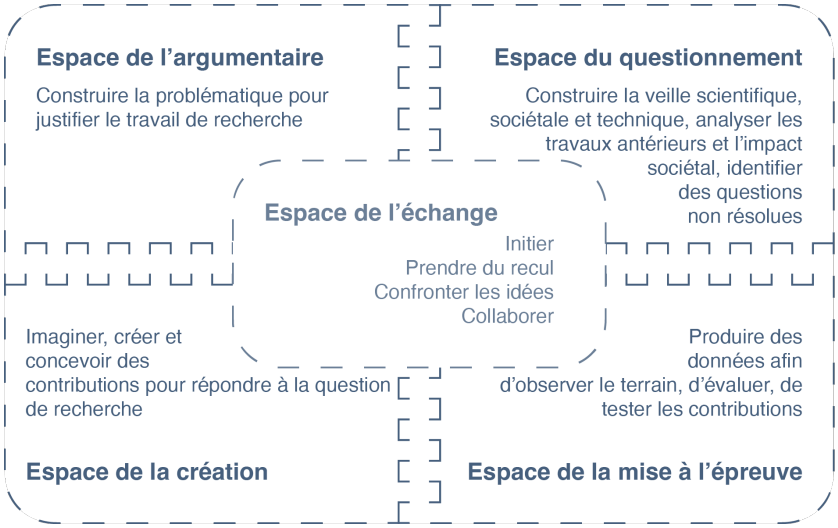


Figure 5. Représentation des 5 espaces de T5S (THEDRE in 5 Spaces)

Bien que T5S n'impose aucun point d'entrée précis, nous préconisons toutefois de commencer le processus dans l'espace de l'échange pour – ce qui est généralement implicite. T5S permet ensuite de naviguer librement et de s'adapter à différentes situations. Voici quelques exemples d'entrée dans T5S :

- Argumentation quand le sujet de thèse et les questions de recherche initiales sont co-construites entre l'encadrement et le doctorant ;
- Questionnement quand le sujet est déjà posé et nécessite du doctorant une acculturation et une réduction du périmètre ;
- Création quand le doctorant a déjà élaboré une solution ou un outil mais qu'ils ne sont pas étayés par la théorie ;
- Mise à l'épreuve quand une étude/observation de terrain amène à se poser une question à problématiser.

3.2 T5S et les autres méthodes

Nous comparons T5S aux méthodes analysées précédemment d'après cinq critères (figure 6) : 1. Objectif : La méthode se positionne comme un outil dédié à

la conception ou à la recherche, 2. Champ lexical : Le vocabulaire utilisé pour décrire la méthode reflète un travail de recherche, 3. Processus : La méthode est décrite comme itérative ou non, 4. Séquentialité : Une séquentialité sous-jacente pilote le processus, 5. La collaboration et l'espace de l'échange sont formalisés.

Critères	HCD	DT	DD	DBR	Mc Kenney & Reeves	ROC	THEDRE	T5S
Objectifs	Conception	Conception	Conception	Recherche et conception	Recherche et conception	Recherche et conception	Recherche et conception	Recherche et conception
Champ lexical adapté recherche	Non	Non	Non	Oui	Oui	Oui, mais le nom des deux cycles est réducteur	Oui	Oui, conçu et évalué en situation
Processus itératif	Oui	Oui	Oui	Un principe	Plusieurs itérations	Double itération	Oui	Non
Séquentialité	Oui	Oui	Oui	.	Oui	Oui	Oui	Non
Formalisation de l'échange	Non	Non	Non	Non	Non	Non	Non	Oui

Figure 6. Comparaisons des méthodes avec T5S

T5S ne s'oppose pas aux méthodes existantes de conduite de la recherche ; elle propose une formalisation en 5 espaces autonomes qui dépasse la vision itérative parfois limitante, en adoptant une approche plus en phase avec le fonctionnement de la recherche, caractérisé par les transitions entre les espaces.

4 Méthode de conception et mise à l'épreuve de T5S

Cette recherche sur T5S s'inscrit dans le constructivisme pragmatique, où la connaissance émerge des interactions entre individus et objets d'étude. Ce cadre valorise l'expérience et les significations contextuelles. Des méthodes qualitatives sont privilégiées pour comprendre les difficultés des doctorants et intégrer leurs besoins dans la conception d'outils pratiques.

T5S est basée sur les concepts de la méthode THEDRE [13], utilisés depuis 2009 dans des formations doctorales (environ 50 doctorants/an). Les retours des doctorants lors de formations sont favorables [4] et ils nous permettent d'améliorer la méthode et son contenu.

En outre, nous avons organisé deux séminaires pour discuter des limites et des contraintes du cycle itératif, constituant la genèse de T5S. Le premier s'est tenu en présentiel avec cinq chercheurs, en mars 2023 et le second s'est tenu en distanciel en septembre 2023 avec huit chercheurs.

La mise à l'épreuve de T5S a évalué : 1) la compréhension des noms et des objectifs proposés pour les cinq espaces, 2) l'utilité et l'utilisabilité de cette organisation modulaire, 3) l'intérêt des allers-retours et non pas d'itération et

[4] Nous avons à notre disposition les résultats des enquêtes réalisées par les écoles doctorales auprès des doctorants

4) l'intérêt de l'espace d'échange. Les 32 participants de cette étude sont des doctorants suivant la formation *Piloter votre thèse* de l'école doctorale de La Rochelle (12) en octobre 2024 et de Grenoble (20) en décembre 2024. Pendant la formation, les doctorants utilisent la méthode T5S et les guides fournis. Ils sont informés que les outils proposés feront l'objet d'une évaluation. Les données qualitatives recueillies : les réponses des participants sur des *post-it*, et les réponses à un questionnaire comprenant 15 questions fermées et deux questions ouvertes. Les résultats reposent sur un comptage des réponses des niveaux d'accord et une analyse des réponses aux questions ouvertes.

5 Résultats

Les retours de la formation à La Rochelle ont conduit à renommer certains espaces : l'"espace du questionnement" devient l'"espace de la fondation" et l'"espace de l'échange" devient l'"espace de la réflexion croisée". Les répondants jugent les noms globalement compréhensibles, avec des variations selon les termes. L'"espace de l'argumentaire" et l'"espace de la création" recueillent un fort consensus (30/32 et 31/32 accords). Cependant, le terme "fondation" est moins clair que "questionnement", et "échange" est jugé plus compréhensible que "réflexion croisée".

Concernant les objectifs, ils sont perçus comme clairs pour la majorité des espaces, notamment pour l'"espace de l'argumentaire" (32/32 accords) et l'"espace de la fondation/questionnement" (30/32). Les espaces de "création" et de "mise à l'épreuve" obtiennent un accord plus modéré. Quelques répondants trouvent les objectifs de la "réflexion croisée" peu clairs. Deux tiers des répondants considèrent le modèle T5S comme utile, flexible, et reflétant les allers-retours du travail de recherche, avec un accord total. Un tiers est plutôt d'accord, mais trois personnes perçoivent moins sa flexibilité. Les commentaires des doctorants confirment ces résultats, soulignant que la recherche est un processus non linéaire et modulaire, permettant des transitions libres entre les phases et des retours en arrière. Comme l'exprime un doctorant : "La progression de la réflexion n'est pas un processus linéaire, de ce fait vouloir cadrer la recherche en phases trop précises est illusoire. Je ressens d'assez près ce fonctionnement modulaire." (D21) Tous les doctorants s'accordent sur le rôle de l'espace de l'échange dans T5S pour encourager la prise de recul ("c'est une structure qui me permet de prendre un peu de recul et améliorer mon travail" (D9)) et favoriser la collaboration ("les [espaces] se complètent malgré tout et surtout l['espace de l'échange] bien mis au centre, de cette façon il boucle le tout dans une réflexion admise et perspicace" (D32)). Ils reconnaissent également son apport en réflexivité ("Apporte de la réflexivité au travail de thèse" (D34)). En revanche, les avis sont plus partagés concernant son utilité pour initier le travail (26/32 accords).

Le modèle T5S offre une démarche de recherche flexible, proche de la réalité, tout en mettant en évidence le rôle clé de l'"espace de l'échange" pour aider les doctorants à prendre du recul et favoriser les interactions scientifiques. Les guides associés aux cinq espaces permettent d'assurer la traçabilité du travail de

recherche. Cependant, les espaces de "création" et d'"échange" nécessitent des ajustements, leurs descriptions dans la littérature étant rares et les guides T5S correspondants insuffisamment précis. En particulier, le guide de l'"espace de la création" semble étroitement lié à la discipline.

6 Conclusion

La méta-méthode T5S offre une approche évolutive pour la recherche, libérée des contraintes des cycles itératifs et des indicateurs d'objectifs. Structurée autour de cinq espaces (Argumentaire, Questionnement, Création, Mise à l'épreuve, Échange), elle favorise une navigation flexible et non linéaire, avec des frontières délibérément poreuses entre les espaces. Les premiers résultats confirment la pertinence de T5S dans l'encadrement des thèses. Toutefois, des études plus approfondies, sur le long terme et en situation réelle, sont nécessaires pour en évaluer pleinement l'efficacité. Bien que la question de la traçabilité, cruciale pour garantir la reproductibilité des recherches "par et pour l'humain", n'ait pas été abordée dans cet article, nos travaux antérieurs ont permis de proposer une méthode et des guides pour y répondre ([13, 14]). Il sera également nécessaire d'étudier si ces guides peuvent contribuer à la traçabilité du processus de recherche avec T5S.

L'espace de l'échange se distingue comme une contribution essentielle, facilitant les transitions entre espaces tout en offrant une perspective nouvelle sur le processus de recherche. En documentant les changements d'espaces et leurs déclencheurs (par exemple, *"Pourquoi passer d'un espace à un autre ?"*), T5S vise à identifier et capitaliser ces éléments déclencheurs pour enrichir les pratiques des communautés scientifiques.

Remerciements. Ce travail est réalisé et financé par le projet ANR TALE4GDA (ANR-23-CE38-0001).

Références

[1] ISO 9241. *ISO 9241-210:2010 - Ergonomics of human-system interaction – Part 210: Human-centred design for interactive systems*. 2010. URL : http://www.iso.org/iso/catalogue_detail.htm?csnumber=52075 (visité le 11/09/2013).

[2] François Baron et Xavier Giloteaux. "La boîte à outils de la sécurité économique". In : *La boîte à outils de la sécurité économique*. Dunod, août 2015, p. 164-167. ISBN : 978-2-10-072846-6. DOI : 10.3917/dunod.moine.2015.01.0164. URL : https://shs.cairn.info/la-boite-a-outils-de-la-securite-economique--9782100728466-page-164?site_lang=fr (visité le 16/10/2024).

[3] Marion Botella et al. "How artists create: Creative process and multivariate factors". In : *Learning and Individual Differences* 26 (2013), p. 161-170. issn : 1041-6080. doi : https://doi.org/10.1016/j.lindif.2013.02.008. url : https://www.sciencedirect.com/science/article/pii/S1041608013000356.

[4] Tim Brown et Barry Katz. "Change by design". In : *Journal of product innovation management* 28.3 (2011). Publisher: Wiley Online Library, p. 381-383.

[5] Richard Buchanan. "Wicked Problems in Design Thinking". In : *Design Issues* 8.2 (1992), p. 5. issn : 07479360. doi : 10.2307/1511637. url : https://www.jstor.org/stable/1511637?origin=crossref (visité le 30/09/2024).

[6] The Design-Based Research Collective. "Design-based research: An emerging paradigm for educational inquiry". In : *Educational Researcher* (2003), p. 5-8. url : http://www.jstor.org/stable/3699927 (visité le 19/04/2016).

[7] The Design Council. *The double Diamond*. url : https://www.designcouncil.org.uk/our-resources/the-double-diamond/ (visité le 01/10/2024).

[8] B. Davis et K. Francis. *Double Diamond Model of Design Thinking*. 2021. url : https://learningdiscourses.com/subdiscourse/double-diamond-model-of-design-thinking/ (visité le 01/10/2024).

[9] W. Edwards Deming. "The new economics. Massachusetts Institute of Technology". In : *Center for Advanced Engineering Study. Cambridge, MA. 240p* (1993).

[10] Interaction Design Foundation. *What is Design Thinking (DT)?* url : https://www.interaction-design.org/ (visité le 01/10/2024).

[11] Philippe Lefebvre. "Comment procèdent réellement les démarches de Design ? Le Double Diamant à l'épreuve". In : *Association Internationale de Recherche en Management Public (AIRMAP)*. Biarritz, France, mai 2018. url : https://minesparis-psl.hal.science/hal-01985188.

[12] Nadine Mandran. "(S')Engager comme doctorant construire une recherche citoyenne". In : *Engagement(s) et citoyenneté(s) numérique(s)*. Sous la dir. de Mines Paris /PSL. Presses des Mines - TRANSVALOR, 2024, p. 65-92. url : https://hal.science/hal-04769885.

[13] Nadine Mandran et Sophie Dupuy-Chessa. "THEDRE: A Traceable Process for High Quality in Human Centred Computer Science Research". In : *International Conference on Information Systems Development (ISD)* (sept. 2017). url : https://aisel.aisnet.org/isd2014/proceedings2017/ISDMethodologies/12.

[14] Nadine Mandran, Mathieu Vermeulen et Estelle Prior. "THEDRE's Framework: Empowering PhD Candidates to Efficiently Implement Design Based-Research". English. In : *Education and Information Technologies* (2022), p. 26.

[15] Susan McKenney et Thomas C. Reeves. *Conducting educational design research*. Routledge, 2018.

[16] J. Nielsen. "Iterative user-interface design". In : *Computer* 26.11 (nov. 1993), p. 32-41. issn : 0018-9162. doi : 10.1109/2.241424. url : http://ieeexplore.ieee.org/document/241424/ (visité le 30/09/2024).

[17] Matthew J Page et al. "The PRISMA 2020 statement: an updated guideline for reporting systematic reviews". In : *BMJ* 372 (2021). doi : 10.1136/bmj.n71. eprint : https://www.bmj.com/content/372/bmj.n71.full.pdf. url : https://www.bmj.com/content/372/bmj.n71.

[18] Éric Sanchez et Réjane Monod-Ansaldi. "Recherche collaborative orientée par la conception: Un paradigme méthodologique pour prendre en compte la complexité des situations d'enseignement-apprentissage". In : *Éducation et didactique* 2 (2015), p. 73-94.

[19] Éric Sanchez et al. "Co. LAB, une plateforme d'accompagnement pour la conception et l'évaluation collaborative de jeux destinés à un usage éducatif". In : *Actes de la 11e Conférence sur les Environnements Informatiques pour l'Apprentissage Humain (EIAH)*. 2023.

[20] Feng Wang et Michael Hannafin. "Design-based research and technology-enhanced learning environments". In : *Educational Technology Research and Development* 53.4 (2005), p. 5-23. url : http://link.springer.com/article/10.1007/BF02504682 (visité le 11/09/2013).

[21] Jingwei Zhu. "On Balancing Flexibility and Compliance of Business Processes: Functional Constraints Modeling and Verification". en. In : *Service-Oriented Computing – ICSOC 2022 Workshops*. Sous la dir. de Javier Troya et al. T. 13821. Series Title: Lecture Notes in Computer Science. Cham : Springer Nature Switzerland, 2023, p. 327-333. doi : 10.1007/978-3-031-26507-5_28. url : https://link.springer.com/10.1007/978-3-031-26507-5_28 (visité le 01/10/2024).

Utiliser un lexique visuel pour mieux accéder à l'activité au sein d'un EIAH

Inès Plessis-Ouzariah
Université de Fribourg, Fribourg, Suisse
ines.plessis-ouzariah@unifr.ch

Trystan Geoffre
Université d'Aix-en-Provence, Aix-en-Provence, France
trystan.geoffre@gmail.com

Résumé. Ce papier propose des analyses générales du lexique visuel (LV), une fonctionnalité d'aide intégrée dans un environnement informatique pour l'apprentissage humain (EIAH), ainsi que les premières observations issues de son expérimentation en classe. Le LV a été utilisé pendant une période de dix semaines dans plusieurs classes afin d'accompagner les élèves dans leurs activités de lecture-compréhension. L'article présente l'outil, son implémentation, et les stratégies d'utilisation identifiées lors de focus groupes menés avec les élèves tout au long de la phase expérimentale. Les résultats qualitatifs obtenus montrent l'intérêt de poursuivre l'analyse de l'utilisation du LV et d'explorer des patterns d'usage dès que les données numériques seront accessibles.

Mots-clés : Fonctionnalités additionnelles, Lexique visuel, Interaction apprenant-outil, Stratégies d'utilisation, Invariants opératoires.

Abstract. This paper presents a general analysis of the visual lexicon (VL), an assistive feature integrated into a computer-based environment for human learning (CBEL), as well as initial observations from its experimentation in the classroom. The LV was used over a ten-week period in several classes to support pupils in their reading and comprehension activities. The article presents the tool, its implementation and the strategies for use identified during focus groups conducted with pupils throughout the experimental phase. The qualitative results obtained show the value of continuing to analyze the use of the LV and exploring patterns of use as soon as the digital data become available.

Keywords: Additional features, Visual Lexicon, Learner-Tool interaction, Usage strategies, Operational Invariants.

1 Introduction

Cet article s'inscrit dans le cadre du travail de doctorat associé au projet financé DOcTA2le-Fr[1] qui vise à développer et mettre à l'épreuve un EIAH ciblant l'enseignement/apprentissage (E/A) de la lecture-compréhension (L-C) en français langue d'enseignement pour le primaire [1, 2]. Outre une approche par la pédagogie du jeu (toutes les activités proposées ciblent des objectifs de lecture-compréhension de textes courts sous la forme de jeux ; des parcours articulant une succession d'activités sont proposés aux élèves), l'EIAH est pensé pour favoriser l'accès de l'apprenant à l'activité, donc à la langue écrite, dans une perspective de conception universelle de l'apprentissage (CUA) [3]. L'interface élève propose ainsi l'accès à différentes fonctionnalités additionnelles (FA) [4] qui peuvent jouer un rôle facilitateur dont un *lexique visuel* qui propose une illustration, une définition et des synonymes du mot recherché (avec entrée phonétique). Nous cherchons ici à comprendre dans quelle mesure les élèves s'approprient cet outil numérique pour soutenir leurs apprentissages, en analysant son usage lors d'une expérimentation en contexte scolaire.

2 Cadre théorique

L'analyse proposée dans cette étude s'inscrit dans une approche mobilisant principalement deux cadres théoriques complémentaires issus de la socio-technique et des sciences de l'éducation.

En premier lieu, l'approche instrumentale développée par Rabardel [5] permet de conceptualiser la transformation d'un artefact en instrument d'apprentissage, grâce aux processus d'instrumentalisation - soit l'appropriation des fonctionnalités techniques par les usagers - et d'instrumentation – c'est-à-dire l'émergence de nouveaux usages à partir de ces fonctionnalités. Cette approche est particulièrement pertinente pour analyser les interactions entre apprenants et dispositifs numériques éducatifs, comme le rappellent Perrin et Draelants [6] dans leurs travaux récents sur l'appropriation des outils numériques en contexte scolaire.

En complément, la théorie des champs conceptuels formulée par Vergnaud [7] offre un cadre d'analyse pour comprendre les schèmes d'action développés par les élèves. Cette théorie met notamment en lumière les invariants opératoires sous-jacents aux pratiques des apprenants, permettant ainsi d'interroger la dynamique de construction des stratégies d'utilisation d'un outil tel que le lexique visuel.

Afin d'enrichir l'analyse de la dimension technique et ergonomique du dispositif, nous mobilisons également des notions issues des travaux de Gomez et Bouchet [8], qui soulignent l'importance de l'utilisabilité pédagogique dans la conception des environnements numériques d'apprentissage.

Enfin, bien que peu de recherches aient spécifiquement étudié l'usage d'un lexique visuel intégré à un EIAH dédié à la lecture-compréhension au primaire, des études

[1] Projet *Documenter et Tester l'Apprentissage Adaptatif à L'École en Français*, **FNS** n°100019_215373

connexes sur des dispositifs proches – tels que les dictionnaires numériques ou les aides lexicales multimodales – mettent en évidence le rôle clé de l'accessibilité et de la personnalisation des outils numériques pour soutenir la compréhension des élèves. À cet égard, les contributions de Carlier et Bétrancourt (2017) [9], ainsi que celles de Sanchez et Baccino (2020) [10], apportent des éclairages utiles sur l'impact de ces dispositifs sur les apprentissages langagiers.

3 La phase expérimentale, les données recueillies, le lexique visuel

Une expérimentation a été menée à l'automne 2024 dans 8 classes (150 élèves ; 2 classes de 3e primaire, 5 classes de 4e primaire, 1 classe de 5e primaire), au cours de 20 séances par classe réparties sur 10 semaines, entre septembre et début décembre) selon un protocole expérimental à cas unique avec lignes de bases multiples [11]: la boite à outils (c'est-à-dire les fonctionnalités additionnelles) était activée de manière aléatoire dans chaque classe, puis désactivée de manière aléatoire à nouveau (tableau 1). On constate que les FA étaient activées (c'est-à-dire disponibles) lors de 82 sessions sur l'ensemble des 160 sessions (8 classes, 20 sessions par classe). La mise à disposition des FA ne signifie pas nécessairement leur utilisation par les élèves ; ils savent que l'outil est accessible et en connaissent les possibilités, mais le recours à leur utilisation effective reste une possibilité. Le but de cette phase expérimentale en classe était donc triple pour documenter :

- si les élèves s'emparent des FA lorsqu'elles sont disponibles (à quel moment ? à quelle fréquence ?) ;
- si les élèves verbalisent des stratégies d'utilisation, et pour quelle efficacité ;
- si l'activation / désactivation des fonctionnalités additionnelles (variable indépendante) a un impact sur les réussites des élèves (variable dépendante).

Table 1. Distribution des phases avec ou sans fonctionnalités additionnelles activées.

Office on the web Frame lasse[1]	Code	N°	S1 16.09 20.09	S2 23.09 27.09	S3 30.09 04.10	S4 07.10 11.10	S5 28.10 01.11	S6 04.11 08.11	S7 11.11 15.11	S8 18.11 22.11	S9 25.11 29.11	S10 02.12 06.12	S11 09.12 13.12	S12 16.12 20.12
	M	1				X[2]								
	CC	2						X						
	AA	3							X					
	K	4								X				
	AM	5	X		Ø						x			
	V	6	X				Ø					x		
	LC	7	X							Ø				
	SA	8	X								Ø			

- [1] L'ordre des classes fait suite à une sélection aléatoire générée par *random.org*, ce paramètre n'est pas négociable.
- [2] La session de jeu dans la semaine (1e session ou 2e session) durant laquelle sont introduites/retirées les fonctionnalités additionnelles fait suite à une sélection aléatoire générées par *random.org*, parmi 2 possibilités (1e ou 2e). Ce paramètre n'est pas négociable.
- X signifie l'introduction de la boîte à outils (fonctionnalités additionnelles).
- Ø signifie le retrait de la boîte à outils (fonctionnalités additionnelles).
- VERT signifie la période durant laquelle la boîte à outils est disponible.
- ROUGE signifie la période durant laquelle la boîte à outils est indisponible.
- ORANGE signifie la période durant laquelle la boîte à outils peut à nouveau être disponible, après un temps de retrait (en faire la demande explicite à

Lors d'une session de travail d'un apprenant, son activité est enregistrée sous forme de traces stockées dans une base de données : ID de l'élève, du parcours, de l'exercice effectué, temps passé, score, activation d'une fonctionnalité additionnelle, laquelle, quand, combien de temps, sur quelle partie de texte, pour lire quels mots ou rechercher quel mot. Nous disposons ainsi du corpus des traces de 150 élèves jouant sur la plateforme et activant ou non les fonctionnalités additionnelles. Il est alors possible de rechercher des invariants d'utilisation et de les lier, aux scores des élèves, à leur progression au fil des sessions, à leurs besoins particuliers (s'ils sont connus), à l'activation pour certains jeux plus spécifiquement, certains textes, phrases ou mots plus spécifiquement, à un moment ou à un autre de l'activité.

Dans cet article, nous étudions plus spécifiquement les données liées à une fonctionnalité additionnelle : le lexique visuel. Pour appréhender l'outil *lexique visuel* (et dans quelle mesure les élèves ont pu l'utiliser, et pourquoi), il est nécessaire de présenter d'abord ses caractéristiques puis de contextualiser son implémentation dans la plateforme. Considéré comme une aide technologique à l'apprentissage (ATA) [12], le lexique visuel interactif est un dictionnaire en ligne illustré et inclusif, conçu pour faciliter la compréhension et l'expression écrite tout en réduisant la charge cognitive liée aux tâches de lecture et d'écriture.



Fig. 1. Interface du lexique visuel

Cet outil propose une entrée phonétique ou orthographique simplifiée (voir **Fig. 1.**) Les définitions fournies par le lexique sont courtes, claires et adaptées à différents publics. Pour enrichir la compréhension, certains mots sont accompagnés de synonymes et d'illustrations visuelles qui visent le renforcement de la mémorisation et la compréhension des termes recherchés.

Le lexique visuel se veut être une solution inclusive : au sein de l'EIAH, il est accessible via un bouton dédié placé sur la zone de jeu. Lors de la phase de jeu, tout utilisateur peut y accéder librement. L'objectif principal est d'offrir aux apprenants la possibilité de rechercher les mots de leurs choix, en fonction de leurs besoins. Les traces d'ouverture / fermeture du lexique visuel, ainsi que du mot recherché, sont enregistrées afin d'analyser ultérieurement, selon une approche qualitative, les motifs de leurs recherches ainsi que l'impact potentiel de cet usage sur leur parcours.

4 Analyse des traces numériques liées au lexique visuel

Nous présentons d'abord quelques éléments concernant les données recueillies, avant de nous centrer sur le lexique visuel et son utilisation.

Au cours de la phase expérimentale, neuf jeux ont progressivement été introduits. Les parcours des élèves étaient ainsi formés d'une succession articulée (et progressive du point de vue didactique) de niveaux de ces neuf jeux. Environ 450 niveaux de jeu ont été sollicités sur l'ensemble des 20 sessions (avec 3 parcours parallèles selon le degré scolaire des classes).
Le **Table 2.** présente le nombre d'activations de chaque fonctionnalité additionnelle (FA) selon les jeux. La synthèse vocale (*playTTS*) est la plus utilisée (5177 activations), suivie du lexique visuel (*dico-img*, 4110) et de la police *opendyslexic* (2058). Le lexique visuel se distingue toutefois par le fait qu'il nécessite une action plus complexe : l'élève doit identifier un mot difficile, le mémoriser puis le saisir manuellement pour accéder à une image et une definition. Il a été mobilisé dans l'ensemble des jeux, mais de manière très inégale. Quatre d'entre eux concentrent plus de 85 % des activations : *Devinette Trésor* (1816), et *Matching* (763), *Par ici ou par là* (483) et *Quand et Pourquoi* (392).

Table 2. Nombre d'utilisations des FA par jeu sur les sessions actives

jeu	class-pad	contrast	dico-img	dyslexic	my-pad	playTTS	space	toolbox-show	total_per_jeu	tot_exo	tot_e1
Les Grams - Chaines	0	0	0	0	0	0	0	0	0	40.0	39.0
Les Grams - Classes	18	0	164	0	11	0	0	63	256	1174.0	1161.0
Les Grams - Silhouettes	3	0	2	0	3	0	0	8	16	339.0	330.0
Les Grams - Substituts	24	0	176	0	14	0	0	245	459	4219.0	3916.0
Les Grams - Tirettes	29	0	73	0	22	0	0	73	197	2256.0	2149.0
[DT] Devinette Tresor	0	0	1816	227	0	888	213	248	3392	5614.0	4494.0
[M] Matching	0	0	763	72	0	888	132	203	2058	10552.0	9560.0
[PDP] Pas de Probleme	0	0	257	242	0	1098	320	238	2155	5875.0	5076.0
[PIOPL] Par ici ou par la	0	5	467	483	0	1309	523	382	3169	8041.0	6836.0
[QEP] Quand et Pourquoi	0	0	392	1034	0	994	1156	1153	4729	6199.0	5397.0
tot_per_type_fa	74	5	4110	2058	50	5177	2344	2613	16431	NaN	0.0

Cette répartition s'explique à la fois par la difficulté lexicale introduite dans certains jeux et par l'absence ou la stylisation du support visuel dans d'autres (*Quand et Pourquoi*, *Par ici ou par là*).

Les figures suivantes présentent l'évolution de l'usage du LV sur plusieurs jeux en observant le nombre d'activations de l'outil (axe bleu) sur le nombre total d'exercices réalisés (axe rouge) au fil des sessions.

Utiliser un lexique visuel pour mieux accéder à l'activité au sein d'un EIAH

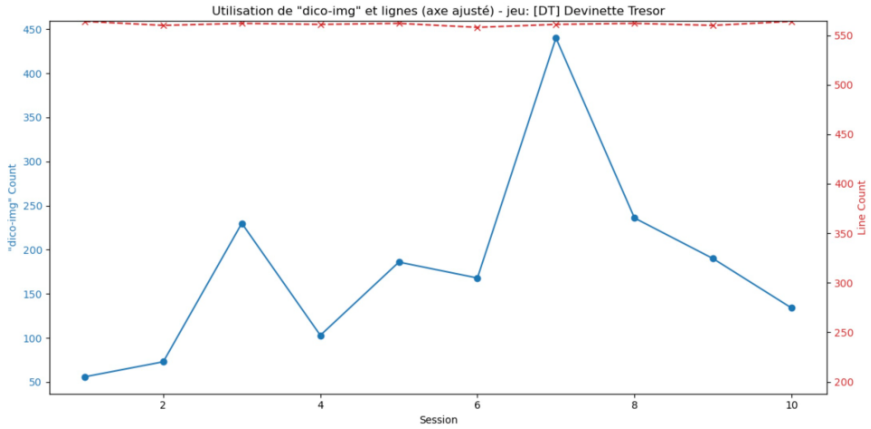


Fig. 2. Evolution du nombre d'activations du LV et du nombre d'exercices réalisés dans le jeu Devinette Trésor

Pour 550 jeux Devinette Trésor par session (en retirant les autres jeux et pour toutes les classes confondues), on observe que le recours au LV augmente fortement jusqu'à la session 7 avant de décroître ensuite. Cela suggère un usage stratégique de cette FA en lien avec la complexité lexicale ou le besoin d'aide dans les premières phases, suivi d'un désengagement probablement une fois les tâches mieux maîtrisées.

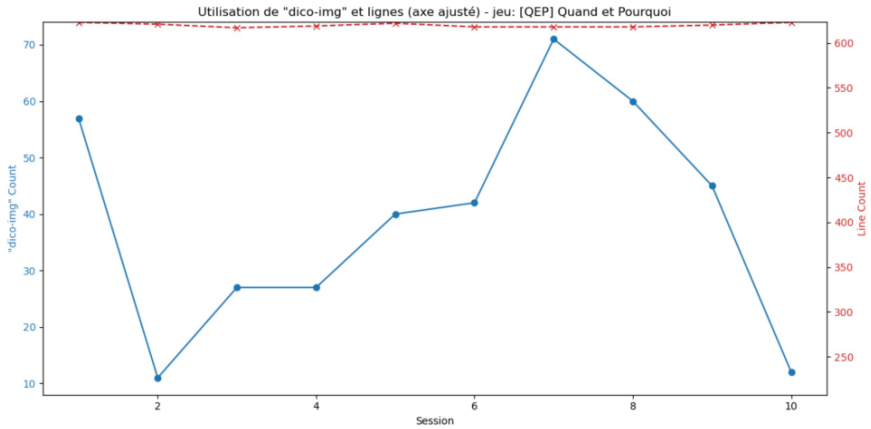


Fig. 3. Evolution du nombre d'activations du LV et du nombre d'exercices réalisés dans le jeu Quand et Pourquoi

En regroupant tous les exercices de Quand et Pourquoi, on observe que le nombre d'activations du LV progresse régulièrement jusqu'à un pic à la session 8, puis décroît. Cette évolution traduit supposément une appropriation progressive de l'outil.

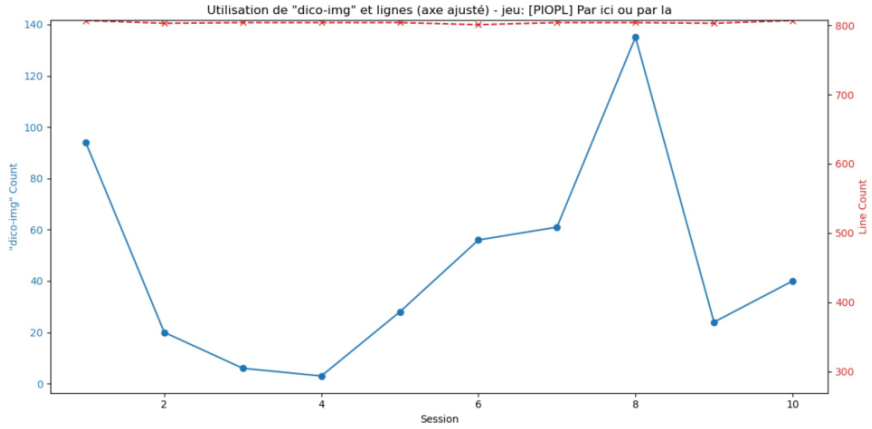


Fig. 4. Evolution du nombre d'activations du LV et du nombre d'exercices réalisés dans le jeu Par ici ou par là

En début de séquence, le nombre d'utilisations du LV est très élevé (presque 100), puis chute fortement pour remonter progressivement. Cette courbe reflète possiblement une phase d'exploration initiale de l'outil, suivie d'un désengagement, puis d'un retour plus ciblé – dynamique typique d'un processus d'appropriation exploratoire.

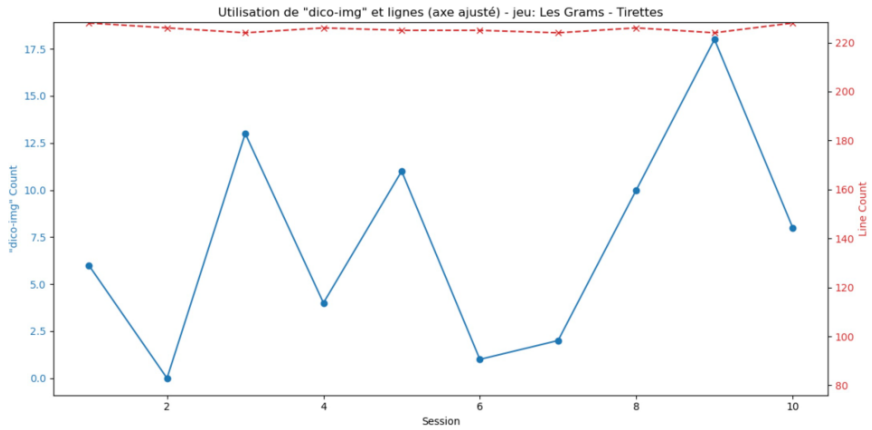


Fig. 5. Evolution du nombre d'activations du LV et du nombre d'exercices réalisés dans le jeu Les Grams - Tirettes

L'usage du LV est irrégulier d'une session à l'autre, sans tendance linéaire, alors que le nombre d'exercices reste stable. Cette instabilité d'usage invite à une analyse plus fine des fonctions mobilisées dans ce jeu.

Tous ces éléments justifient l'ouverture vers une analyse qualitative centrée sur l'expérience des élèves.

Utiliser un lexique visuel pour mieux accéder à l'activité au sein d'un EIAH

5 L'expérience d'utilisation du point de vue élève

Pour compléter le corpus de données (cf. partie **4**), une approche qualitative a été privilégiée : des focus groupes (FG) [13] menés dans 3 classes de 4e primaire d'une même école. Le protocole des focus groupes (FG) visait à approfondir l'analyse de l'usage du LV en contexte de jeu. Les interactions des élèves (clics, activation du LV, mots recherchés) étaient suivies en temps réel. À l'issue de chaque session avec FA activées, les élèves utilisateurs du LV (2 à 13 participants) étaient invités à un FG, dont les objectifs leur étaient explicités en amont : explorer les motivations, perceptions et stratégies d'usage. Les échanges, guidés par des questions ouvertes, ont permis de documenter les dimensions individuelles et collectives de l'appropriation de l'outil. Chaque entretien, d'une durée moyenne de 8 minutes, a été enregistré et intégralement transcrit en vue de l'analyse. La fiabilité des analyses repose donc sur le croisement entre données qualitatives et quantitatives.

5.1 La perception de l'outil : réflexions globales des élèves

La perception initiale d'un outil joue un rôle déterminant dans sa pré-appropriation (Vaujany [14]). Deux éléments principaux relatifs aux attentes et habitudes d'apprentissage des élèves influencent leur usage du LV :

1. L'intégration d'illustrations a été perçue positivement : pour la majorité des élèves, elles facilitent la compréhension en associant visuel et texte. Cependant, certains soulignent que l'image seule ne suffit pas toujours, notamment pour les mots abstraits ou les verbes, parfois absents du corpus illustré ;
2. Certains élèves ont indiqué utiliser le LV non pour l'image, mais pour obtenir des informations grammaticales (classe du mot), notamment dans le cadre du jeu des Grams – Tirettes. Ici, l'image reste secondaire au profit de la fonctionnalité de consultation rapide, dans un objectif ciblé, ce qui concorde avec la ponctualité de son usage (cf. **Fig. 5).**

De façon générale, les élèves valorisent la présence conjointe d'une image et d'une définition, qui permet de croiser deux modes de compréhension. L'outil est ainsi perçu comme un soutien pertinent à la lecture mais son usage réel dépend des stratégies adoptées en situation.

5.2 Analyse des stratégies d'utilisation à partir des cadres théoriques

L'analyse des transcriptions a permis d'identifier des schèmes d'action récurrents – au sens de la théorie des champs conceptuels de Vergnaud [7] – qui révèlent des logiques différenciées d'appropriation du LV. En mobilisant également l'approche instrumentale [5], deux grandes familles de stratégies peuvent être distinguées : celles relevant d'une instrumentalisation *fonctionnelle* de l'outil, intégrée à l'activité

d'apprentissage, et celles traduisant une instrumentalisation *exploratoire*, souvent encore détachée des objectifs de la tâche.

Stratégies d'Utilisation Fonctionnelle

Nous appelons stratégies fonctionnelles les usages orientés vers la résolution immédiate d'un problème ou l'exécution d'une tâche spécifique. Elles s'ancrent dans des schèmes construits autour d'invariants opératoires tels que : *"je rencontre une difficulté → j'active le LV pour la résoudre"*. Trois sous-types principaux sont identifiés :

• Compréhension lexicale, chercher le sens d'un mot inconnu pour comprendre une consigne ou un texte ;

• Résolution d'un blocage, utiliser le LV comme appui dans un moment de difficulté ;

• Recherche grammaticale, consulter la classe du mot pour progresser dans des jeux ciblés.

Ce type de logique est illustré par exemple dans *Devinette Trésor* (cf. **Fig. 2.**) et *Quand et Pourquoi* (cf. **Fig. 3.**), où l'on observe une montée progressive des activations du LV dans un contexte de jeux à forte composante lexicale. Ces usages traduisent une **instrumentation réussie** : l'élève a intégré le LV comme ressource mobilisable dans l'activité. La recherche est motivée par une intention claire, souvent liée à l'efficacité et à la réussite immédiate.

Stratégies Exploratoires

Ces stratégies relèvent d'une instrumentalisation en cours : l'élève interagit avec l'outil de manière plus libre, sans objectif pédagogique explicite. Elles s'appuient sur des invariants opératoires moins stabilisés, du type : *"je clique pour voir"* ou *"j'essaie pour tester"*. Deux types principaux sont observés :

• Curiosité spontanée, consultation de mots par envie ou intérêt, sans lien direct avec la tâche, ce qui suggère une nécessité de relance pour une utilisation efficiente : être attiré par une fonctionnalité illustre un intérêt en surface pour l'outil si l'élève ne va pas plus loin dans son utilisation. Une relance (par exemple de l'enseignant) pourrait permettre de dépasser l'exploration pour s'approprier l'aide proposée à dessein.

• Utilisation accidentelle ou de diversion, activation par ennui, erreur ou distraction. Ces comportements sont également observés lors de l'exploration d'autres fonctionnalités de la plateforme.

Les stratégies exploratoires traduisent une démarche plus ouverte, elles révèlent un désir d'explorer les fonctionnalités de l'outil sans objectif précis. Ce type de comportement peut être mis en regard des usages observés dans Par ici ou par là (cf. **Fig. 4.**), où l'on constate une forte sollicitation initiale suivie d'un désengagement, typique d'une phase d'exploration non consolidée. De manière plus ponctuelle, Les Grams – Tirettes (cf**. Fig. 5.**) reflète une instrumentalisation spécifique mais irrégulière, parfois non orientée vers la résolution directe de tâche. Ce type d'usage témoigne d'une

instrumentalisation moins efficace, dans laquelle l'élève navigue dans l'outil sans nécessairement l'utiliser pour son apprentissage. Les invariants opératoires associés aux stratégies exploratoires incluent donc un processus d'essais et d'erreurs, ainsi qu'une démarche autonome d'exploration des ressources disponibles. Ces recherches ne sont motivées ni par des objectifs d'efficacité ni de bonne résolution de la tâche.

Il semble donc que les stratégies fonctionnelles et exploratoires reflètent des besoins différents chez les élèves : d'une part, une recherche orientée vers la résolution de problèmes et, d'autre part, une utilisation détournée ou ludique de l'outil.

Ces deux familles de stratégies trouvent un écho dans les dynamiques d'usage observées dans les données de traces. Les figures présentées précédemment illustrent plusieurs profils d'appropriation du LV selon les jeux : certaines activités, comme *Devinette Trésor* ou *Quand et Pourquoi*, révèlent une mobilisation croissante du LV, suggérant un ancrage progressif dans les routines de résolution. D'autres, comme *Par ici ou par là* ou *Les Grams – Tirettes*, montrent des usages plus ponctuels ou irréguliers, en lien avec des stratégies exploratoires ou des objectifs spécifiques. Ce croisement entre données quantitatives et qualitatives renforce l'interprétation des comportements observés comme relevant de schèmes différenciés d'usage, allant de l'exploration libre à une instrumentalisation pleinement intégrée dans l'activité.

6 Conclusion

Ce travail ouvre ainsi la voie à une réflexion plus large sur l'appropriation progressive des EIAH comme instruments cognitifs, construite à l'intersection des artefacts, des activités ludiques et des apprentissages. En effet, cette première étude sur l'usage du lexique visuel dans l'EIAH GamesHub met en lumière deux formes d'appropriation : une instrumentalisation exploratoire, souvent non finalisée, et une instrumentation fonctionnelle intégrée à la tâche. Le croisement entre données de traces et discours d'élèves révèle des schèmes d'action différenciés, en lien avec la nature des jeux et les profils d'élèves. L'outil est globalement perçu comme utile, indépendamment du niveau en lecture-compréhension. Reste à interroger la stabilité des usages dans le temps, et leur diffusion auprès d'élèves initialement moins engagés.

La prochaine phase expérimentale (janvier–avril 2025), fondée sur un protocole à cas unique avec lignes de base multiples (11 classes, 20 sessions), permettra d'observer l'évolution de ces stratégies dans un contexte d'*adaptive learning*. En ajustant les parcours aux besoins des élèves, on peut faire l'hypothèse d'un glissement vers des usages plus ciblés et stratégiques du LV.

Références

1. Karoui, A., Alvarez, L., Geoffre, T., Guin, N., Lefevre, M., Lachand-Pascal, V., Ramalho, M.: Towards an Automated Adaptive Learning Web Platform Through Personalization of Language Learning Pathways. In: Hilliger, I., Muñoz-Merino, P. J., De Laet, T., Ortega-Arranz, A., Farrell, T. (eds) Educating for a New Future: Making Sense of Technology-

Enhanced Learning Adoption. *Lecture Notes in Computer Science*, vol. 13450, pp. 448–454. Springer, Cham (2022). DOI: 10.1007/978-3-031-16290-9_35
2. Alvarez, L., Karoui, A., Ramalho, M., Geoffre, T.: Fondements pédago-didactiques d'un environnement informatique d'apprentissage destiné aux jeunes apprenants de l'école primaire. In: Proc. 10e Conférence sur les Environnements Informatiques pour l'Apprentissage Humain (EIAH'2021)
3. CAST: What is UDL? *UDL Resource*. https://www.udlresource.com/udl.html
4. Rodi, M., Dherbey Chapuis, N., Geoffre, T., Alvarez, L.: Faciliter l'accès à l'apprentissage du français pour tous au sein de parcours adaptables de jeux éducatifs numériques. *Alsic* 25(1) (2022).
5. Rabardel, P.: Les hommes et les technologies: approche cognitive des instruments contemporains. Armand Colin, Paris (1995).
6. Perrin, N., Draelants, H.: L'appropriation des outils numériques en éducation: une lecture par l'approche instrumentale. *Revue Française de Pédagogie* (2021).
7. Vergnaud, G.: Concepts et schèmes dans une théorie opératoire de la représentation. *Psychologie Française* 30(3), 254–251 (1985).
8. Gomez, M., Bouchet, F.: L'utilisabilité pédagogique des environnements numériques d'apprentissage. *Revue STICEF* (2018).
9. Carlier, S., Bétrancourt, M.: Impact des aides multimodales sur la compréhension en lecture. *Canadian Journal of Learning and Technology* 43(1) (2017).
10. Sanchez, E., Baccino, T.: Les environnements numériques pour la lecture: état de la recherche. *Revue des Sciences de* l'Éducation 46(2) (2020).
11. Référence anonymisée pour la révision
12. Ferron, J.M., Moeyaert, M., Van den Noortgate, W., Beretvas, S.N.: Estimating causal effects from multiple-baseline studies: Implications for design and analysis. *Psychological Methods* 19, 493–510 (2014). https://doi.org/10.1037/a0037038.
13. Morgan, D.L.: *Focus Groups as Qualitative Research*. 2nd edn. SAGE Publications, Thousand Oaks (1997).
14. De Vaujany, F.: Pour une théorie de l'appropriation des outils de gestion: vers un dépassement de l'opposition conception-usage. *Management & Avenir* 9(3), 109–126 (2006). https://doi.org/10.3917/mav.009.0109.

Étude comparative entre réfugiés ukrainiens et apprenants en situation de guerre sur la créativité dans des tâches de robotique pédagogique

Margarida Romero[1,2], Oksana Strutynska [2]

[1] Universitat Internacional de Catalunya, Barcelone, Espagne
margarida.romero@gmail.com
[2] Université Côte d'Azur, Nice, France
oksana.strutynska@univ-cotedazur.fr

Résumé. La résolution créative de problèmes (RCP) est influencée par l'état émotionnel de l'apprenant-joueur. Cette étude examine l'impact de la guerre, ainsi que de l'exil qu'elle entraîne, en tant que situations stressantes susceptibles d'affecter la pensée divergente mobilisée en RCP. Deux groupes ont été impliqués : des réfugiés ukrainiens vivant en France et des élèves résidant à Kyïv, en Ukraine. À travers la tâche CreaCube, nous comparons les performances des deux groupes afin de mieux comprendre l'influence du stress sur la pensée divergente. Les résultats montrent que les participants de Kyïv présentent une flexibilité et une originalité supérieures à celles des réfugiés. Ces résultats suggèrent qu'un stress chronique, vécu en situation de guerre, pourrait être associé à une capacité accrue de génération d'idées nouvelles dans un contexte de robotique pédagogique.

Mots-clés : Créativité, Pensée divergente, Fluidité, Flexibilité, Originalité, Robotique pédagogique.

Abstract. Creative problem solving (CPS) is influenced by the emotional state of the learner-player. This study examines the impact of war, as well as the exile it causes, as stressful situations likely to affect the divergent thinking involved in CPS. Two groups were involved: Ukrainian refugees living in France and students residing in Kyiv, Ukraine. Through the CreaCube task, we compare the performance of both groups to better understand the influence of stress on divergent thinking. The results show that the participants from Kyiv demonstrate greater flexibility and originality than the refugees. These findings suggest that chronic stress, experienced in a war situation, may be associated with an increased ability to generate new ideas in an educational robotics context.

Keywords: Creativity, Divergent Thinking, Fluency, Flexibility, Originality, Educational Robotics.

1 Introduction

La robotique éducative connaît un essor important, porté par la popularité croissante des robots et l'élargissement de leurs domaines d'application [1, 2]. Dans ce cadre, la créativité peut être stimulée à travers la construction de configurations variées, analysées selon les dimensions de la pensée divergente : fluidité, flexibilité et originalité. Toutefois, le développement de la créativité requiert un contexte émotionnel favorable. Qu'il soit individuel ou collectif, un climat positif est essentiel pour soutenir la résolution créative de problèmes (RCP) [5]. Cependant, les recherches sur l'impact du stress sur la créativité révèlent une relation ambivalente, sujette à controverse : le stress peut affecter la créativité tant positivement que négativement, selon les circonstances. Étant un vécu subjectif, ses effets varient selon sa nature. Certaines situations, telles que la guerre, sont considérées comme parmi les plus stressantes pour les enfants [6] et relèvent d'un stress chronique. Plusieurs études suggèrent que plus le stress est intense, plus les productions créatives peuvent être riches et diversifiées. À la lumière de ces travaux, il apparaît nécessaire d'examiner l'impact de situations hautement stressantes, comme la guerre, sur la RCP.

2 La pensée divergente dans la résolution de problèmes

Dans cette étude, nous nous intéressons aux tâches de résolution créative de problèmes (RCP), qui mobilisent simultanément la pensée divergente (génération d'idées) et la pensée convergente (sélection et raffinement), selon un processus cyclique et itératif. La pensée divergente, définie comme la capacité à produire un grand nombre d'idées ou de solutions originales à un problème donné, joue un rôle central dans la RCP [7]. Elle se décline en trois composantes majeures, identifiées par Guilford [8] : la *fluidité* (quantité d'idées produites), la *flexibilité* (diversité des types de réponses) et l'*originalité* (caractère rare ou inattendu des idées dans un contexte donné). Dans le cadre de cette recherche, nous analysons ces trois composantes à travers les productions concrètes d'élèves réalisées avec des cubes robotiques modulaires. La tâche proposée s'inscrit dans un contexte pédagogique en robotique éducative et a été menée auprès de deux groupes d'apprenants : des élèves vivant à Kyïv, en Ukraine, et des réfugiés ukrainiens résidant en France. Cette comparaison vise à explorer comment des contextes de vie contrastés influencent la pensée divergente en situation de RCP.

3 Méthodologie

1.1 Participants

L'expérience a impliqué 33 participants ukrainiens, répartis en deux groupes : des réfugiés ukrainiens en France (n=17) et des participants de Kyïv (n=16). Les participants ont un âge moyen de 10,65 ans. Les participants à Kyïv vivent des situations stressantes liées aux conditions de guerre. Les participants ont pris part à une activité ludique avec des robots modulaires pour évaluer leur pensée divergente. Lors des ateliers organisés en France, nous avons travaillé avec des familles issues de la

communauté de réfugiés ukrainiens. A Kyïv, les participants sont invités à la Faculté de Mathématiques, Informatique et Physique de l'Université d'État Dragomanov (Ukraine), en respectant les mesures de sécurité en temps de guerre. Les participants ont rapporté avoir apprécié les activités ludiques. Pendant l'atelier, toutes les mesures nécessaires ont été prises pour assurer la sécurité des participants. En cas de sirènes aériennes, les participants descendent dans l'abri avec la chercheure à Kyïv. Une fois le danger écarté, l'activité de jeu en robotique pédagogique est reprise une fois que les élèves se sentent à l'aise pour jouer.



Fig. 1. Élève ukrainien réfugié en France en train de résoudre la tâche CreaCube.

1.2 Procédure pour l'évaluation de la pensée divergente

Pour évaluer la pensée divergente, les participants ont réalisé la tâche CreaCube, une activité de résolution créative de problèmes (RCP) basée sur l'assemblage de cubes robotiques modulaires. L'objectif était de concevoir un véhicule autonome capable de se déplacer d'un point de départ à une destination [9, 10]. Cette tâche permet d'analyser trois composantes de la pensée divergente, selon le cadre du Test des usages alternatifs (*Alternate Use task*, AUT) de Guilford [11] : la fluidité (nombre d'idées), la flexibilité (nombre de catégories d'idées) et l'originalité (idées rares, produites par moins de 5 % des pairs). Chaque participant a effectué deux essais (A1 et A2). Nous analysons dans cette étude la pensée divergente du premier essai (A1). Les productions ont été enregistrées en vidéo, puis analysées à l'aide d'un logiciel intégré à la plateforme CreaMaker. Ce système, basé sur une base de configurations préenregistrées, permet d'attribuer des scores automatisés pour les trois composantes de la pensée divergente.

2 Résultats

Les vidéos de la première instance (A1) de la tâche CreaCube ont été analysées à l'aide de la plateforme CreaMaker, permettant de calculer automatiquement les scores de fluidité, flexibilité et originalité. Les résultats ont ensuite été traités avec le logiciel JASP (v. 0.17.1). Les élèves de Kyïv obtiennent des scores significativement plus élevés en flexibilité (m = 2,18 ; écart-type = 1,97) et en originalité (m = 1,19 ; écart-type = 1,47) par rapport aux réfugiés ukrainiens en France. La figure 3 illustre cette différence pour l'originalité : les élèves de Kyïv présentent une moyenne de 1,18 (écart-type = 1,47), contre 0,13 (écart-type = 0,35) pour ceux en France. Le test t de Welch

confirme une différence significative entre les deux groupes pour l'originalité ($t(16,821) = 2,784$, $p = 0,013$).

3 Discussion

L'analyse comparative entre les élèves réfugiés ukrainiens en France et ceux vivant à Kyïv met en évidence des différences significatives, en particulier sur la composante d'originalité lors de la première instance (A1) de la tâche CreaCube. Les scores plus élevés du groupe de Kyïv peuvent s'expliquer par leur exposition quotidienne à des situations complexes et imprévisibles liées au contexte de guerre : coupures d'électricité, pénuries d'eau ou de chauffage, fermeture des commerces, pannes d'équipements, interruptions régulières des activités scolaires ou sociales en raison des alertes aériennes, etc. Ce contexte génère un stress chronique qui, paradoxalement, peut stimuler la créativité, notamment par la nécessité constante de s'adapter et de trouver des solutions originales à des problèmes concrets. Les résultats suggèrent ainsi que le stress contextuel vécu par les élèves à Kyïv peut favoriser certaines dimensions de la pensée divergente, en particulier l'originalité. Cette capacité à générer des solutions inédites dans des situations nouvelles pourrait être reconnue et valorisée dans les pratiques éducatives, notamment comme manifestation de résilience. En outre, la participation à une activité ludique et interactive comme CreaCube semble avoir eu un effet positif sur l'engagement des élèves de Kyïv, leur offrant un moment de répit et de plaisir cognitif dans un environnement fortement contraint. Cette activité, tout en sollicitant la créativité, permet également d'introduire la robotique pédagogique de manière tangible et accessible, même en contexte de crise.

Financement

ANR CreaMaker (ANR-18-CE38-0001) et le programme PAUSE du Collège de France.

Références

1. Rapti, S., Sapounidis, T.: "Critical thinking, Communication, Collaboration, Creativity in kindergarten with Educational Robotics": A scoping review (2012–2023). *Comput. Educ.* 210, 104968 (2024). https://doi.org/10.1016/j.compedu.2023.104968
2. Chevalier, M., Giang, F., Boissin, A., Iordanou, S., Krasniqi, A., Bertin, F.: The role of feedback and guidance as intervention methods to foster computational thinking in educational robotics learning activities for primary school. *Comput. Educ.* 180, 104431 (2022). https://doi.org/10.1016/j.compedu.2022.104431
3. Leroy, A., Romero, M., Cassone, L.: Interactivity and materiality matter in creativity: educational robotics for the assessment of divergent thinking. *Interact. Learn. Environ.* 1–12 (2021). https://doi.org/10.1080/10494820.2021.1875005
4. Romero, M., Freiman, V., Rafalska, M.: Techno-creative Problem-Solving (TCPS) Framework for Transversal Epistemological and Didactical Positions: The Case Studies of CreaCube and the Tower of Hanoi. In: Michelsen, C. et al. (eds.) *Mathematics and Its Connections to the Arts and Sciences*, pp. 245–274. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-10518-0_13

5. Stinkeste, C., Napala, A., Romero, M.: Impact of Team Climate on Co-Creativity in STEAM Education in Primary Education. *Creative Education* 12(8), 1977–1994 (2021). https://doi.org/10.4236/ce.2021.128151
6. Palace, M., Zamazii, O., Terbeck, S., Bokszczanin, A., Berezovski, T., Gurbisz, D., Szwejka, L.: Mapping the factors behind ongoing war stress in Ukraine-based young civilian adults. *Appl. Psychol. Health Well-Being* 16(3), 868–885 (2024).
7. Runco, M.A.: *Problem Finding, Problem Solving, and Creativity*. Greenwood Publishing Group, Westport (1994).
8. Guilford, J.: Creativity: Yesterday, today and tomorrow. *J. Creative Behav.* 1, 3–14 (1967).
9. Romero, M., DeBlois, L., Pavel, A.: Créacube, comparaison de la résolution créative de problèmes, chez des enfants et des adultes, par le biais d'une tâche de robotique modulaire. *MathémaTICE* 61 (2018). http://revue.sesamath.net/spip.php?article1104#
10. Romero, M., DeBlois, L.: Analyse du processus de construction de connaissances dans des activités de programmation à l'école. *Can. J. Sci. Math. Techn. Educ.* 22(2), 405–421 (2022). https://doi.org/10.1007/s42330-022-00210-9

La perception d'utilité d'outils d'échafaudage numériques pour soutenir la résolution de problèmes : comparaison entre trois contextes d'application en enseignement supérieur

Chantal Tremblay[1]

[1] Université du Québec à Montréal, Montréal, QC H3C 3P8, Canada
Tremblay.chantal@uqam.ca

Résumé. Cette étude porte sur la conception d'outils d'échafaudages numériques (OÉN) pour soutenir l'apprentissage de la résolution de problèmes complexes auprès d'étudiants de 1er cycle universitaire en gestion. Elle vise à compa-rer les perceptions d'étudiants ayant participé à l'une des trois itérations de cette étude à l'égard de l'utilité et la facilité d'utilisation de ces outils. Ces OÉN cor-respondent principalement à des questionnements cognitifs, stratégiques et mé-tacognitifs. Les outils cognitifs et stratégiques seraient généralement perçus utiles, tandis que l'utilité des outils métacognitifs serait peu perçue. La formation initiale, la convivialité des outils et leur caractère agréables pourraient influencer la perception de facilité d'utilisation et, conséquemment, l'usage des outils. Ainsi, il est proposé de poursuivre la recherche sur la conception d'OÉN en s'as-surant que les étudiants soient mieux informés de leur pertinence pour soutenir leurs apprentissages et en choisissant une application numérique conviviale pour les intégrer, comme un outil d'intelligence artificielle générative.

Mots-clés : Outils d'échafaudage numériques, résolution de problèmes complexes, recherche orientée par la conception, modèle d'acceptation technologique, enseignement supérieur

Abstract. This study focuses on the design of digital scaffolding tools to support complex problem-solving learning among undergraduate management students. It aims to compare the perceptions of students who participated in one of the three iterations of the study regarding the usefulness and ease of use of these tools. These tools primarily consist of cognitive, strategic, and metacognitive prompts. Cognitive and strategic tools are generally perceived as useful, whereas the usefulness of metacognitive tools is less commonly recognized. Initial train-ing, objective usability, and their perceived enjoyment may influence students' perceptions of ease of use, and consequently, their actual use of the tools. There-fore, it is recommended that research on the design of digital scaffolding tools continue, ensuring that students are better informed about their relevance in sup-porting learning and that user-friendly digital platforms—such as generative ar-tificial intelligence tools—are selected for their integration.

Keywords: digital scaffolds, complex problem solving, design-based research, technology acceptance model, higher education

1 Contexte : Développer la compétence de résolution de problèmes complexes (RPC) d'étudiants en gestion

L'essor des technologies et de l'intelligence artificielle générative transforme le marché du travail, en modifiant la nature des tâches que doivent effectuer les travailleurs [39]. À cet effet, plusieurs référentiels de compétences du 21e siècle [32, 34] exposent les compétences essentielles que les étudiants universitaires devraient développer durant leur formation. Parmi elles, on retrouve la résolution de problèmes complexes (RPC), qui constitue également une compétence professionnelle que doivent développer les étudiants de plusieurs écoles de gestion [1, 11]. Toutefois, des études ont montré des écarts entre les compétences de diplômés récents en gestion et les attentes des gestionnaires [1, 16]. Par exemple, certains étudiants éprouvent des difficultés à produire une solution adéquate, ce qui peut être causé par une incapacité à se représenter adéquatement le problème ou à déterminer l'ensemble des conséquences d'une solution [33]. Un manque de connaissances disciplinaires [9] ou des compétences métacognitives sous-développées pourraient être la cause de ces difficultés [13], ce qui limiterait leur capacité à se construire une démarche de RPC et à exploiter les contenus disciplinaires pour élaborer des solutions adéquates [5].

Ce projet visait donc à soutenir l'apprentissage de la RPC d'étudiants universitaires en gestion, à l'aide d'outils numériques intégrés à une application numérique (Karuta) conçus en s'appuyant sur la théorie de l'étayage de Bruner [4]. Ces outils prennent la forme de questionnements, de conseils d'experts ou de rétroaction, visant à soutenir la compréhension du problème, le développement d'une solution argumentée et l'autoévaluation. La méthodologie s'inspire de recherches orientées par la conception, qui visent le développement d'activités pédagogiques pour résoudre des problématiques complexes en éducation [21]. Ces recherches sont conçues pour être itératives et appliquées en contexte authentiques, où praticiens et chercheurs collaborent à chaque étape [21, 27, 29]. Dans ce projet, trois itérations ont eu lieu avec trois groupes d'enseignants qui avaient identifié des lacunes chez leurs étudiants lors de la RPC. S'inspirant du modèle de McKenney et Reeves [21], chaque itération s'est déroulée en trois phases, dont la première visait l'analyse des lacunes des étudiants pour en déterminer les causes (section 2). La seconde phase correspondait au design de l'activité, puis à la construction d'outils et ressources permettant sa mise en œuvre (section 3). La troisième phase consistait au déploiement et à l'évaluation de l'activité selon les objectifs de la recherche (sections 4 et 5). L'évaluation permet de formuler des recommandations pour concevoir des outils numériques perçus utiles qui incitent à les utiliser (section 6).

2 Analyse de la situation problème : les lacunes des étudiants pour chaque étape de la RPC

En gestion, la RPC suit généralement une séquence de cinq étapes (**Fig. 1**). La comparaison entre la façon dont des experts (professionnels) solutionnent des problèmes complexes et des novices (étudiants) permet d'exposer certaines lacunes de ces derniers.

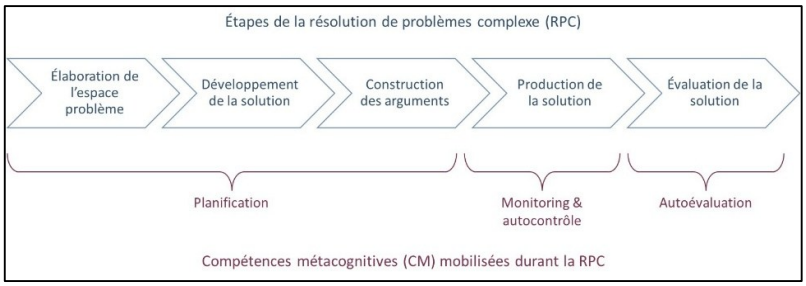


Fig. 1. Étapes de la résolution de problèmes complexes en gestion

D'abord, l'étudiant élabore son espace problème en tentant de se représenter adéquatement le problème, en comprendre les causes, les parties prenantes et son objectif [12, 22, 37]. Les novices accordent parfois moins d'importance et de temps à cette représentation que les experts [5, 30], ce qui limite leur capacité à comprendre le problème, à en repérer les éléments importants, à saisir son objectif ou à y associer des contenus disciplinaires pertinents [9, 17, 31]. Ensuite, l'étudiant développe sa solution : il va formuler plusieurs solutions, évaluer chacune d'elles puis choisir la plus adéquate, compte-tenu de l'objectif [12]. Les problèmes en gestion n'ont généralement pas de solution unique : l'étudiant devra construire des arguments pour justifier la pertinence de sa solution. Les novices n'évaluent pas toujours adéquatement l'ensemble des conséquences de leur solution, ce qui réduit leur capacité à rédiger des arguments cohérents et convaincants pour la soutenir [31, 37]. Après ses premières étapes, l'étudiant réalise une production pour exposer sa solution au problème. En gestion, cette production peut correspondre à un texte argumentatif, un rapport d'analyse ou un discours justifiant la pertinence de recommandations. Pour soutenir son apprentissage, l'étudiant devrait terminer la RPC en autoévaluant sa solution et sa démarche [9].

Tout au long de la RPC, l'étudiant mobilise ses connaissances disciplinaires et ses compétences métacognitives [9, 26]. Celles-ci correspondent à l'ensemble des compétences qui permettent de s'autoréguler et de contrôler ses processus cognitifs [20]. La planification est essentielle pour comprendre le problème et choisir les actions à exécuter pour élaborer une solution adéquate [9, 26]. Le monitoring et l'autocontrôle permettent de réguler ses processus cognitifs et ses comportements, pour conserver son attention sur l'objectif du problème et d'agir pour le résoudre [22]. L'autoévaluation permet de juger sa solution et sa démarche pour développer ses habiletés de RPC. La littérature soutient que les experts mobiliseraient plus souvent et de façon plus efficaces leurs compétences métacognitives que les novices, notamment l'autoévaluation [5, 28].

3 Design de l'activité : mobilisation de la théorie de l'étayage (scaffolding) de Bruner pour concevoir les outils intégrés à l'application numérique

Pour soutenir le processus de RPC d'étudiants en gestion, ce projet visait à concevoir des outils d'échafaudage numériques (OÉN). L'intérêt de ces outils est motivé par des recherches, méta-analyses et recensions d'écrits qui montrent que les OÉN peuvent soutenir l'apprentissage en présence [3, 15], en formation à distance en ligne [8] et l'apprentissage autorégulé [7, 38, 41]. Des études montrent que les OÉN influencent positivement l'apprentissage de la RPC [9, 10, 13], mais peu ont été menées en gestion. Ces OÉN sont basés sur la théorie de l'étayage de Bruner (1983), définie par l'assistance fournie par un tuteur à un apprenti, afin qu'il puisse accomplir une tâche qu'il ne peut réaliser seul [4, 39]. Selon Bruner, l'apprenti sera capable de réaliser la tâche par lui-même lorsqu'il aura internalisé le dialogue avec son tuteur. Le rôle du tuteur consiste à évaluer le niveau de son apprenti à intervalles réguliers, afin d'ajuster et de retirer progressivement son assistance, qui prend la forme d'un dialogue axé sur le questionnement, les conseils et la rétroaction [4]. L'étayage contribuerait à l'apprentissage de la RPC grâce au potentiel du dialogue pour expliciter des processus implicites [2], permettant à l'étudiant de se représenter chaque étape de la RPC et les tâches associées, puis de les internaliser pour se construire une démarche de RPC [4, 25]. L'explicitation contribuerait aussi au développement de compétences métacognitives.

En enseignement supérieur, plusieurs recherches substituent le tuteur par des OÉN [3, 8, 15, 19]. En général, ces outils sont intégrés à une application numérique et prennent la forme de questionnements, de conseils d'experts et de rétroactions [41]. Bien que les termes *digital scaffolding*, *digital scaffolds* ou *computer-based scaffolding* soient utilisés pour désigner une variété d'outils numériques dans la littérature, nous considérons que ce qui permet de qualifier un outil d'OÉN est sa cohérence avec la théorie de Bruner. Un OÉN doit inciter au dialogue intérieur et viser l'internalisation. Il peut assister l'étudiant en structurant la réalisation de tâches complexes, si cette structuration est explicitée, visible et ainsi suscite son internalisation [14, 18]. Les OÉN sont habituellement catégorisés selon leur intention pédagogique : soutenir l'utilisation adéquate des contenus ou la représentation du problème (OÉN cognitifs), structurer la démarche et approfondir le raisonnement (OÉN stratégiques), susciter la mobilisation de compétences métacognitives (OÉN métacognitifs), soutenir la motivation (OÉN motivationnels) ou faciliter l'usage d'outils numériques (OÉN procéduraux) [3, 41].

Dans ce projet, trois expérimentations d'OÉN intégrés à l'application Karuta ont eu lieu. La première a eu lieu dans un cours d'économie où les étudiants devaient prendre position sur un enjeu d'économie publique. La seconde fut dans un cours de comptabilité dans lequel les étudiants collaboraient à construire un outil de mesure de la performance. La troisième s'est déroulée dans un cours préparant des équipes d'étudiants à un concours interuniversitaire de résolution de cas en gestion (disciplines variées). Les OÉN ont été adaptés pour chaque contexte, bien que les étapes de RPC demeuraient les mêmes. Ils ont été améliorés lors des 2e et 3e itérations en considérant la rétroaction reçue par les étudiants lors de chaque collecte de données. Pour chacun, des OÉN

cognitifs, stratégiques et métacognitifs ont été créés en collaboration avec la chercheuse principale et les enseignants. Les OÉN cognitifs correspondaient à des questionnements à chacune des trois premières étapes de la RPC. Des OÉN stratégiques ont été ajoutés à l'étape de la construction des arguments pour amener les étudiants à identifier les limites de leurs solutions choisies et les améliorer. Aussi, l'application a été conceptualisée de manière à inciter les étudiants à suivre une démarche structurée de RPC, notamment par l'ajout d'onglets et d'encadrés suggérant les étapes à réaliser (OÉN stratégiques) qui ont été révisés après chaque itération. Des tableaux et des encadrés et des couleurs ont été ajoutés pour inciter davantage à leur utilisation. Les OÉN métacognitifs correspondaient à des questions de planification du travail et d'autoévaluation. Une liste de vérification et des conseils d'experts ont été ajoutés lors de la 1re itération, mais leur faible perception d'utilité a mené à leur abandon ensuite. Les questions d'autoévaluation ont été rendues obligatoire lors de la 2e itération pour susciter leur usage, car l'utilité était faiblement perçue lors de la 1re itération. Une rétroaction standardisée fournie après la remise de la solution (OÉN cognitifs) a été ajoutée à la 1re itération, qui a été complémentée par une rétroaction de l'enseignant à la 2e itération. Le Tableau 1 présente une synthèse des OÉN pour chaque itération. Le `matériel supplémentaire` montre des captures d'écran et détaille leur contenu.

Tableau 1. Description des outils d'échafaudage numériques selon les itérations

Étape de la RPC	Itération 1 (économie)	Itération 2 (comptabilité)	Itération 3 (gestion – variées)
Espace problème	3Q cognitives 1Q métacognitive	3Q cognitives 1Q métacognitive	6Q cognitives 2Q métacognitives
Développement solution	3Q cognitives	4Q cognitives	3Q cognitives 2C métacognitifs
Construction arguments	3Q stratégiques	4Q stratégiques	2Q stratégiques
Production solution	5 items métacognitifs avec conseils experts, rétroaction standardisée	rétroaction standardisée et personnalisée	Sans objet
Autoévaluation solution	1 vidéo d'expert, 4Q métacognitives	4Q (ind.) métacognitives	3Q (collabo), 5Q (ind.) métacognitives

Légende : Q = Question, C = Conseil

4 Déploiement de l'activité : Objectifs, méthodes de collecte et d'analyse des données pour cette contribution

Cette communication vise à présenter des résultats préliminaires découlant d'un objectif spécifique du projet, qui correspond à décrire l'intention d'utilisation des OÉN, en s'appuyant sur le modèle Technology Acceptance Model (TAM) de Davis *et al.* [6] et ses adaptations TAM2 [36] et TAM3 [35]. Ce modèle montre que les perceptions d'utilité et de facilité d'utilisation d'un outil numérique sont les principaux déterminants de

l'intention d'utilisation, qui correspond à une attitude favorable ou défavorable prédisposant à l'utilisation réelle de l'outil. Notre objectif est donc d'expliquer la perception d'utilité des différents OÉN en s'appuyant sur ses déterminants identifiés dans le modèle TAM2 et de décrire les perceptions de facilité d'utilisation des OÉN en mobilisant des déterminants du modèle TAM3 (**Erreur ! Source du renvoi introuvable.**).

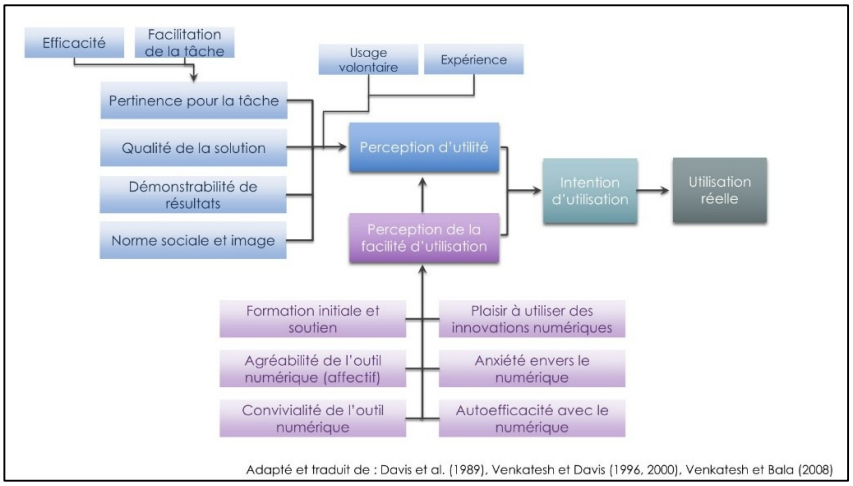


Fig. 2. Technology Acceptance Model [6] et intégration des facteurs TAM2 [36] et TAM3 [35]

Les données ont été collectés lors d'entretiens individuels et de groupe semi-dirigés (Tableau 2). Le guide d'entretien contenait une section portant sur l'intention d'utilisation des outils, dans laquelle des questions étaient posées pour discuter principalement de la perception d'utilité des différents outils et de l'application Karuta en général. Les questions portaient sur la perception de pertinence des outils et les retombées de leur usage sur l'efficacité, la facilitation de la tâche et la qualité de la solution produite. Puis, la perception de facilité d'utilisation était abordée et complétée par des questions de relance sur la convivialité et l'agréabilité des outils, ainsi que la formation reçue. Les participants qui ont expliqué ne pas avoir utilisé certains outils étaient invités à en expliquer les raisons. Des questions de relance visaient à déterminer si des déterminants de la perception d'utilité ou de facilité d'utilisation en étaient la cause. À noter que les résultats de la première itération ont été publiés dans un article en 2022 : cette communication porte sur une synthèse des trois itérations pour comparer les perceptions et expliquer ce qui influence l'intention d'utilisation d'OÉN conçus pour soutenir la RPC.

Les verbatim d'entretiens ont été analysés en s'appuyant sur une démarche d'analyse thématique [24]. L'arbre de codage (disponible en `matériel supplémentaire`) a été élaboré en s'appuyant sur le modèle TAM, tout en permettant d'ajouter d'autres codes lorsque nécessaire. Le premier niveau correspond aux perceptions d'utilité et de facilité d'utilisation des OÉN. Le deuxième niveau permet d'étiqueter les discours du participant selon les déterminants des modèles TAM2 et TAM3. Pour la perception

d'utilité seulement, un troisième niveau distingue le type d'OÉN auquel le participant se réfère lorsqu'il le précise ou s'il tient un discours sur l'ensemble des OÉN ou sur l'application Karuta. Enfin, le dernier niveau contient les codes qui permettent de scinder les discours selon que la perception du participant est favorable, défavorable ou neutre à l'égard de l'utilité ou de la facilité d'utilisation des OÉN.

Tableau 2. Participants à la collecte de données

Entretiens	Itérations		
	1 (hiver 2020)	2 (hiver 2024)	3 (automne 2024)
Individuels	n = 6	n = 4	n = 6
De groupe	Gr. 1 : n = 9 Gr. 2 : n = 6	Gr. 1: n = 4	Sans objet

5 Évaluation de l'activité: résultats préliminaires

5.1 Une perception d'utilité des OÉN qui témoigne d'un besoin de soutien lors de la RPC

Les résultats indiquent que les étudiants perçoivent généralement utiles les questions cognitives soutenant l'élaboration de leur espace problème, notamment en les amenant à réfléchir à des éléments qu'ils n'auraient pas considérés autrement. Par exemple, un participant (2e itération) soutient que ces questions « étaient pertinentes parce qu'elles nous amenaient à réfléchir profondément (…), ça permettait d'aller plus en profondeur en fait ». Toutefois, l'utilité des questions métacognitives liées à l'espace problème ont peu été utilisées et peu considérées utiles, et ce, aux trois itérations.

La perception d'utilité des questions associées au développement de la solution est mitigée entre les itérations. À la première, une seule participante avait indiqué avoir répondu aux questions de cette section pour un problème qu'elle trouvait plus difficile. Les autres ont mentionné qu'ils avaient déjà choisi leur solution après avoir élaboré leur espace problème. À l'opposé, les participants de la 2e itération exposent une perception d'utilité de ces questions globalement favorable, principalement à cause de la structuration en deux parties. Comme les problèmes étaient résolus en équipe, les étudiants devaient d'abord répondre à ces questions et à des questions stratégiques visant la construction des arguments de façon individuelle. Lorsque l'équipe se réunissait, les réponses de chaque membre s'affichaient dans l'application et permettaient alors la confrontation des points de vue, ce qui a été particulièrement apprécié et jugé utile par les participants du groupe de discussion. À la troisième itération, certains participants ont apprécié le format « tableau » des questions de développement de la solution et de construction des arguments, qui leur permettaient de mieux se représenter les avantages et les limites de chaque solution.

Néanmoins, la majorité des participants ont indiqué qu'ils n'ont pas mobilisés les questions métacognitives (incluant les items métacognitifs et les conseils d'experts), car ils n'en voyaient pas la pertinence ni l'utilité. Plusieurs ne comprenaient pas l'utilité des questions d'autoévaluation de la solution. Comme la rétroaction standardisée était perçue peu utile à la 1re itération, elle a été complémentée par une rétroaction personnalisée par l'enseignant lors de la 2e itération, ce qui a été plus ou moins perçu utile. À la troisième itération, cette fonctionnalité était disponible, mais elle n'a pas été utilisée. Dans les trois cas, les étudiants ont préféré recevoir une rétroaction à l'oral de leur enseignant qui leur semblait plus précise et détaillée.

Enfin, certains participants ont considéré utile la « structure » de l'application, car cela guidait leur démarche, facilitait leur processus notamment en les concentrant sur la tâche à effectuer et les rendait ainsi plus efficaces.

.

5.2 Une perception de facilité d'utilisation mitigée

Les participants de la première et troisième itération considèrent presque tous que les OÉN étaient faciles à utiliser grâce à leur convivialité, et ce, malgré un manque de formation initiale perçu à la première itération. Pour eux, la perception de facilité d'utilisation n'était pas déterminante de leur perception d'utilité. Lors de la deuxième itération, des participants ont mentionné avoir éprouvé certaines difficultés à utiliser les outils, notamment par manque de formation initiale, ce qui a parfois contribué à réduire leur utilisation et leur perception d'utilité. Ils ont indiqué que l'application n'était pas autant conviviale, intuitive ou agréable à utiliser que des applications commerciales, réduisant ainsi leur perception de facilité d'utilisation. Dans l'ensemble, les participants se considèrent assez compétents avec le numérique et cela semble peu influencer leur perception de facilité d'utilisation.

6 Discussion et conclusion

Ces résultats préliminaires indiquent que les questions cognitives et stratégiques visant l'élaboration de l'espace problème, le développement de la solution et la construction des arguments développés dans ce projet constitueraient une forme d'assistance perçue utile et pertinente par la majorité des participants. Ce résultat expose leur besoin d'être mieux accompagnés pour la RPC, ce qui suggère qu'ils sont plus novices qu'experts [37] et explique les écarts observés lorsqu'ils entrent sur le marché du travail [1]. D'ailleurs, leur absence de compréhension de l'utilité des questions d'autoévaluation témoigne de lacunes de compétences métacognitives, pourtant essentielles à la RPC [9].

Ces constats justifient la pertinence de concevoir de tels outils en misant sur leur perception d'utilité pour convaincre les étudiants de les exploiter. En effet, nos résultats soutiennent que ce déterminant est le principal facteur qui mène à une intention d'utilisation favorable des outils, bien que les déterminants de la convivialité et du caractère agréable des outils (facilité d'utilisation) soient à considérer pour susciter leur usage.

Certaines limites de ce projet pourraient orienter la recherche future. À l'instar de [38], nos résultats soulignent l'importance du rôle de l'enseignant pour présenter les outils afin d'en exposer leur pertinence et susciter leur intention d'utilisation. Le

manque de convivialité amène à réfléchir au choix de l'application dans lequel ils sont intégrés. De fait, l'application Karuta comporte certaines limites qui pouvaient rendre l'utilisation plus ardue, surtout pour la RPC collaborative (itération 2). Une piste à considérer consisterait à configurer un outil d'intelligence artificielle générative qui dialoguerait avec l'étudiant en suivant la théorie de l'étayage. Cela pourrait améliorer les perceptions de convivialité (facilité d'utilisation), mais surtout la perception d'utilité grâce à une production de questionnements et de rétroactions personnalisées, haussant ainsi la perception de pertinence et d'amélioration de la qualité de la solution produite.

Remerciements. Cette étude a reçu un soutien financier du Conseil de recherche en sciences humaines du Canada et du Fonds de recherche du Québec – Société et culture.

Déclaration de conflits d'intérêts. L'autrice déclare n'avoir aucun conflit d'intérêt.

Références

1. AACSB. (2018). AACSB Industry Brief: Lifelong Learning and Talent Management. The Association to Advance Collegiate Schools of Business (AACSB).
2. Belland, B. R., Walker, A. E., Olsen, M. W. et Leary, H. (2015). A Pilot Meta-Analysis of Computer-Based Scaffolding in STEM Education. Journal of Educational Technology & Society, 18(1), 183 197.
3. Belland, B. R., Walker, A. E., Kim, N. J. et Lefler, M. (2017). Synthesizing Results From Empirical Research on Computer-Based Scaffolding in STEM Education: A Meta-Analysis. Review of Educational Research, 87(2), 309 344. https://doi.org/10.3102/0034654316670999
4. Bruner, J. S. (1983). Le développement de l'enfant: savoir faire, savoir dire. Presses universitaires de France.
5. Bruning, R. H., Schraw, G. J. et Ronning, R. R. (2011). Chapter 8 Problem Solving and Critical Thinking. Dans Cognitive Psychology and Instruction (5th edition). Pearson.
6. Davis, F. D., Bagozzi, R. P. et Warshaw, P. R. (1989). User Acceptance of Computer Technology: A Comparison of Two Theoretical Models. Management Science, 35(8), 982-1003. https://doi.org/10.1287/mnsc.35.8.982
7. Devolder, A., Braak, J. van et Tondeur, J. (2012). Supporting self-regulated learning in computer-based learning environments: systematic review of effects of scaffolding in the domain of science education. Journal of Computer Assisted Learning, 28(6), 557 573. https://doi.org/10.1111/j.1365-2729.2011.00476.x
8. Doo, M. Y., Bonk, C. et Heo, H. (2020). A Meta-Analysis of Scaffolding Effects in Online Learning in Higher Education. International Review of Research in Open and Distributed Learning, 21(3), 60 80. https://doi.org/10.19173/irrodl.v21i3.4638
9. Ge, X. et Land, S. M. (2004). A Conceptual Framework for Scaffolding Ill-Structured Problem-Solving Processes Using Question Prompts and Peer Interactions. Educational Technology Research and Development, 52(2), 5 22. https:// /doi/10.1007/BF02504836
10. Ge, X., Chen, C.-H. et Davis, K. A. (2005). Scaffolding Novice Instructional Designers' Problem-Solving Processes Using Question Prompts in a Web-Based Learning Environment. Journal of Educational Computing Research, 33(2), 219 248. https://doi.org/10.2190/5F6J-HHVF-2U2B-8T3G
11. HEC Montréal. (2018). Objectifs du B.A.A. et vision des finissants. http://www.hec.ca/etudiants/mon-programme/baa/vision-objectifs/index.html

12. Jonassen, D. H. (2011). Learning to solve problems: a handbook for designing problem-solving learning environments. Routledge.
13. Kauffman, D. F., Ge, X., Xie, K. et Chen, C.-H. (2008). Prompting in Web-Based Environments: Supporting Self-Monitoring and Problem Solving Skills in College Students. Journal of Educational Computing Research, 38(2), 115 137.
14. Kim, M. C. et Hannafin, M. J. (2011). Scaffolding Problem Solving in Technology-Enhanced Learning Environments (TELEs): Bridging Research and Theory with Practice. Computers & Education, 56(2), 403 417. https://doi.org/10.1016/j.compedu.2010.08.024
15. Kim, N. J., Belland, B. R. et Walker, A. E. (2018). Effectiveness of Computer-Based Scaffolding in the Context of Problem-Based Learning for Stem Education: Bayesian Meta-analysis. Educational Psychology Review, 30(2), 397 429. https://doi.org/10.1007/s10648-017-9419-1
16. Koys, D. J., Thompson, K. R., Martin, W. "Marty" et Lewis, P. (2019). Build it and they will come: Designing management curricula to meet career needs. Journal of Education for Business, 94(8), 503-511. https://doi.org/10.1080/08832323.2019.1580244
17. Lacombe, N. (2010). Le jugement professionnel : conceptions d'étudiants universitaires en sciences comptables à partir d'une analyse qualitative d'entrevues de groupe [Mémoire de maîtrise, Université Laval].
18. Lin, X., Hmelo, C., Kinzer, C. K. et Secules, T. J. (1999). Designing technology to support reflection. Educational Technology Research and Development, 47(3), 43 62. https://doi.org/10.1007/BF02299633
19. Martha, A. S. D., Santoso, H. B., Junus, K. et Suhartanto, H. (2019). A Scaffolding Design for Pedagogical Agents within the Higher-Education Context. Dans Proceedings of the 2019 11th International Conference on Education Technology and Computers (p. 139 143). Association for Computing Machinery. https://doi.org/10.1145/3369255.3369267
20. Mayer, R. E. (1998). Cognitive, metacognitive, and motivational aspects of problem solving. Instructional Science, 26(1 2), 49 63. https://doi.org/10.1023/A:1003088013286
21. McKenney, S. et Reeves, T. C. (2018). Conducting Educational Design Research. Routledge. https://doi.org/10.4324/9781315105642
22. Nelson, T. O. et Narens, L. (1990). Metamemory: A Theoretical Framework and New Findings. Psychology of Learning and Motivation, 26, 125 173. https://doi.org/10.1016/S0079-7421(08)60053-5
23. Newell, A. et Simon, H. A. (1972). Human problem solving. Prentice-Hall.
24. Paillé, P. et Mucchielli, A. (2016). Chapitre 11. L'analyse thématique. Dans L'analyse qualitative en sciences humaines et sociales (Quatrième édition, p. 235-312). Armand Colin. http://catalogue.bnf.fr/ark:/12148/cb45070160r
25. Palincsar, A. S. (1986). The Role of Dialogue in Providing Scaffolded Instruction. Educational Psychologist, 21(1 2), 73 98. https://doi.org/10.1080/00461520.1986.9653025
26. Poissant, H., Poellhuber, B. et Falardeau, M. (1994). Résolution de problèmes, autorégulation et apprentissage. Canadian Journal of Education / Revue canadienne de l'éducation, 19(1), 30 44. https://doi.org/10.2307/1495305
27. Reeves, T. C. (2006). Design research from a technology perspective. Dans J. Van den Akker, K. Gravemeijer, S. McKenney et N. Nieveen, Educational Design Research. Taylor & Francis Group. http://ebookcentral.proquest.com/lib/umontreal-ebooks/detail.action?docID=274476
28. Schunk, D. H. (2014). 7. Cognitive Learning Processes. Dans D. H. Schunk, Learning theories: an educational perspective (Sixth edition, p. 230-279). Pearson.

29. Stockless, A. (2016). Le processus d'adoption d'une innovation pédagogique avec les TIC par les enseignants [Thèse de doctorat inédite, Université de Montréal]. Papyrus. https://hdl.handle.net/1866/18373
30. Swanson, H. L., O'Connor, J. E. et Cooney, J. B. (1990). An Information Processing Analysis of Expert and Novice Teachers' Problem Solving. American Educational Research Journal, 27(3), 533 556. https://doi.org/10.3102/00028312027003533
31. Tawfik, A. A., Kim, K. et Kim, D. (2020). Effects of case library recommendation system on problem solving and knowledge structure development. Educational Technology Research and Development, 68(3), 1329-1353. https://doi.org/10.1007/s11423-020-09737-w
32. Tremblay, C. et Poellhuber, B. (2022). Analyse qualitative de référentiels de compétences du XXIe siècle, numériques et informationnelles : tendances mondiales observées. Formation et profession, 30(2), 1. https://doi.org/10.18162/fp.2022.648
33. Tremblay, C., Poellhuber, B. et Kozanitis, A. (2022). Les outils d'échafaudage numériques pour l'apprentissage de la résolution de problèmes complexes : analyse des déterminants de l'intention d'utilisation. Revue internationale des technologies en pédagogie universitaire, 19(3), 16-43. https://doi.org/10.18162/ritpu-2022-v19n3-02
34. van Laar, E., van Deursen, A. J. A. M., van Dijk, J. A. G. M. et de Haan, J. (2020). Determinants of 21st-Century Skills and 21st-Century Digital Skills for Workers: A Systematic Literature Review. SAGE Open, 10(1), 2158244019900176. https://doi.org/10.1177/2158244019900176
35. Venkatesh, V. et Bala, H. (2008). Technology Acceptance Model 3 and a Research Agenda on Interventions. Decision Sciences, 39(2), 273-315. https://doi.org/10.1111/j.1540-5915.2008.00192.x
36. Venkatesh, V. et Davis, F. D. (2000). A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies. Management Science, 46(2), 186-204. https://doi.org/10.1287/mnsc.46.2.186.11926
37. Voss, J. F., Greene, T. R., Post, T. A. et Penner, B. C. (1983). Problem-Solving Skill in the Social Sciences. Psychology of Learning and Motivation, 17, 165-213. https://doi.org/10.1016/S0079-7421(08)60099-7
38. Wong, J., Baars, M., Davis, D., Zee, T. V. D., Houben, G.-J. et Paas, F. (2019). Supporting Self-Regulated Learning in Online Learning Environments and MOOCs: A Systematic Review. International Journal of Human–Computer Interaction, 35(4 5), 356 373. https://doi.org/10.1080/10447318.2018.1543084
39. Wood, D., Bruner, J. S. et Ross, G. (1976). The role of tutoring in problem solving. Journal of child psychology and psychiatry, and allied disciplines, 17(2), 89 100. https://doi.org/10.1111/j.1469-7610.1976.tb00381.x
40. World Economic Forum. (2023). The Future of Jobs Report 2023. World Economic Forum. https://www.weforum.org/reports/the-future-of-jobs-report-2023/
41. Zheng, L. (2016). The effectiveness of self-regulated learning scaffolds on academic performance in computer-based learning environments: a meta-analysis. Asia Pacific Education Review, 17(2), 187 202. https://doi.org/10.1007/s12564-016-9426-9Author, F.: Article title. Journal 2(5), 99–110 (2016)

EIAH
2025 Lille
Environnements Informatiques pour l'Apprentissage Humain

Posters et démonstrations

Identifier un environnement numérique propice à l'expérience optimale d'apprentissage : prémices d'une revue de littérature

Marjolaine Chagrot[1], Anthony Cherbonnier[1], Jérôme Guérin[2], Jean Heutte[1]

[1] Université de Lille, ULR 4354 - CIREL
marjolaine.chagrot@univ-lille.fr
anthony.cherbonnier@univ-lille.fr
jean.heutte@univ-lille.fr
[3] Université de Bretagne Occidentale, CREAD (EA 3875)
jerome.guerin@univ-brest.fr

Résumé. La digitalisation de la formation et les nouvelles technologies ouvrent de nouvelles perspectives pour améliorer l'expérience d'apprentissage des apprenants. L'objectif de cette revue systématique est d'identifier les environnements numériques favorables à l'émergence du flow, conceptualisé par Csikzentmihalyi, et leurs caractéristiques. Une recherche dans cinq bases de données, en suivant la méthode PRISMA, a permis de sélectionner 126 publications sur le sujet, qui devront être analysées par la suite.

Mots-clés : flow, autotélisme-flow, environnement numérique d'apprentissage, revue de littérature, e-formation

Abstract. The digitization of training and new technologies are opening up new prospects for improving the learning experience of learners. The aim of this systematic review is to identify digital environments conducive to the emergence of flow, as conceptualized by Csikzentmihalyi, and their characteristics. A search of six databases, using the PRISMA method, resulted in the selection of 126 publications on the subject, which will be analyzed in turn.

Keywords: flow, digital learning environment, literature review, digitization, learning experience

1 Introduction

Dans un contexte de transition numérique et de digitalisation de la société, l'enseignement se transforme pour utiliser au mieux ces technologies. Dans l'enseignement supérieur, les universités proposent de plus en plus de typologies différentes de dispositifs de formation, incluant de nouvelles modalités et pratiques. Avec la e-formation, on assiste à une multiplication des environnements potentiels d'apprentissage puisque la formation à distance n'a par définition, pas de lieu qui lui soit attaché. Cette ouverture des dispositifs laisse une belle part d'autonomie aux apprenants. Suite à cette évolution majeure du formalisme des formations, les objets de recherche en science de l'éducation se sont déplacés vers une recherche interdisciplinaire, se mêlant aux sciences des technologies de l'information et de la communication, mais aussi à la psychologie, en s'intéressant particulièrement aux théories agentives (e.g. autodétermination, autorégulation, auto-efficacité) pour mieux comprendre l'apprenant et les défis auxquels il doit faire face, ainsi qu'à la psychologie positive pour l'aider à les surmonter.

En s'inscrivant dans le courant des technologies positives pour l'apprentissage décrit par Molinari et al. (2021), notre démarche sur dans cette recherche va être de comprendre comment favoriser l'expérience optimale d'apprentissage en environnement numérique. Pour cela, nous allons réaliser une revue de littérature systématique, avec la méthode PRISMA.

2 Cadres théoriques

2.1 Flow

L'expérience optimale, aussi appelé "flow" ou "autotélisme-flow" est défini comme : "*un état d'épanouissement lié à une profonde implication et au sentiment d'absorption que les personnes ressentent lorsqu'elles sont confrontées à des tâches dont les exigences sont élevées et qu'elles perçoivent que leurs compétences leur permettent de relever ces défis. Le flow est décrit comme une expérience optimale au cours de laquelle les personnes sont profondément motivées à persister dans leur activité.* [...]" (EFRN, 2014). Il s'agit d'un état dynamique de bien-être, de plénitude, de joie, d'implication totale (Heutte, 2019, p.163), conceptualisé par Csikszentmihalyi (1975).

Cette expérience de bien-être intense peut être observé dans diverses activités : le travail et les études, l'apprentissage (e.g. Peterson & Miller, 2004 ; Rathunde & Csikszentmihalyi, 2005), l'enseignement (Coleman, 2014), les loisirs sportifs, culturels et créatifs, les activités de recherche, les jeux vidéo (e.g. Ivory & Magee, 2009, Thin et al. 2011), le médical... Une récente revue de littérature recense 252 études réalisées sur le flow entre 2000 et 2016 et démontre que les contextes investigués sont vastes (Peifer et al., 2022). Elle révèle également que la plupart de ces recherches concernent le niveau individuel, étudié sous différents prismes : la personnalité, la motivation, les émotions, la cognition, le comportement, ou la physiologie. En Technologie de l'Information et de la Communication (TIC), les travaux suggèrent que dans l'expérience globale

avec la technologie, des concepts tels que l'enjouement et le flow sont des variables ayant un rôle potentiellement importants dans les théories d'acceptation de l'usage des technologies (Agarwal & Karahanna, 2000).

En science de l'éducation, interroger le flow des apprenants en formation revient à s'intéresser au rôle du bien-être dans les résultats de l'apprentissage, et dans le développement de compétences. Nakamura & Csíkszentmihályi (2009) ont recensé de multiples études illustrant les avantages du flow dans différents contextes éducatifs notamment l'augmentation de l'engagement scolaire, de la motivation, du développement des compétences, de l'amélioration des performances et de l'estime de soi, et une diminution de l'anxiété. D'après Heutte (2011), le flow en contexte éducatif peut-être modélisé selon quatre dimensions : (1) contrôle de l'action, (2) absence de préoccupation à propos du soi (dilatation de l'ego), (3) altération de la perception du temps, (4) expérience autotélique (ou intrinsèque : bien-être procuré par la tâche elle-même).

La communauté scientifique reconnaît un "caractère prometteur des recherches concernant le flow dans les environnements numériques" (Choi et al. 2007). Cependant, Heutte (2019, p.178) observe que "le champ de la e-formation des adultes n'est que très peu investigué, de plus la plupart de ces travaux sont relativement récents".

2.2 Environnements numériques

Comme l'explique Charlier (2019), pour comprendre l'Environnement Numérique d'Apprentissage (ENA) en tant qu'objet de recherche, il ne suffit pas de rechercher dans la littérature les références renvoyant directement aux recherches sur les ENA. Après un travail mené sur les terminologies associées à ces notions et pour guider de futures recherches, Charlier propose de définir les ENA de la manière suivante : "*l'ENA est un système d'apprentissage ouvert intégrant les acteurs et leurs instruments numériques et non numériques. Ce système constitue potentiellement le lieu d'apprentissage émergent. La conception et la régulation de l'ENA peuvent être plus ou moins contrôlées : totalement par l'apprenant, négociées ou complètement contrôlées par l'institution. […] "* (2019, p.63).

Les objets de recherche concernant les environnements numériques en contexte éducatif se concentrent sur l'apport des médias dans les dispositifs de formation. On trouve également beaucoup de recherches s'intéressant aux interactions entre pairs sur les plateformes d'apprentissage. Ces travaux cherchent en partie à répondre aux difficultés d'ordre pédagogique, didactique ou technique (Chevry Pebayle & Rossini, 2017) constatées des apprenants dans les formations médiatisées par ordinateur. Nombre de ces études développent des modèles, méthodes et outils afin de permettre la mise en place d'aides réactives et proactives dans ces dispositifs de formation à distance.

3 Méthodologie

Dans le but d'identifier les environnements numériques d'apprentissage soutenant un état de flow, il est nécessaire de réaliser un état de l'art idéalement inspiré de méthodes documentées, explicites et reproductibles. Cette revue de littérature va permettre de faire un point sur l'état des connaissances et des études effectuées sur le flow en environnement numérique en formation, pour ensuite décrire, analyser et caractériser ces environnements. Notre question de recherche (QR) est la suivante :

- **QR1 :** Est-ce qu'un environnement numérique permet de faire émerger un état de flow ?

Nous avons établi, pour guider notre analyse, deux sous-questions (SQRs) :

- **SQR1 :** Quels sont les instruments de mesure du flow en environnement numérique ?
- **SQR2 :** Quelles sont les caractéristiques des environnements numériques propices à l'émergence du flow ?

3.1 Méthode

Afin d'assurer la rigueur méthodologique et la transparence de notre revue, nous avons mobilisé la méthode *Preferred Reporting Items for Systematic reviews and Meta-Analyses* (PRISMA, Page et al., 2021). Cette approche nous permet d'expliciter les différentes étapes du processus de revue systématique - de l'identification des sources à leur inclusion finale - en garantissant la reproductibilité et la traçabilité du travail effectué.

3.2 Stratégie de recherche

Notre revue systématique cherche à identifier les études réalisées autour des environnements numériques d'apprentissage qui soutiennent le flow des apprenants. Notre recherche croise trois concepts : le flow, les environnements numériques et le contexte éducatif. Notre requête est construite avec les fonctions booléenne « AND » et « OR » afin d'élargir nos résultats à un maximum de contenus pertinents :

("flow experience*" OR "flow theory*" OR "flow state" OR "optimal experience*" OR "challenge-skill balance" OR "Csikszentmihalyi" OR "team flow" OR "collective flow" OR "shared flow" OR "social flow" OR "EduFlow") AND ("software" OR "online tool*" OR "online environment" OR "numerical" OR "numerical tool*" OR "digital tool*" OR "digital environment" OR "electronic tool*" OR "computer based

tool*" OR "interactive environment" OR "computer assisted" OR "computer-aided" OR "web-based" OR "computer-supported" OR "technology-supported") AND ("train*" OR "educat*" OR "learn*" OR "instruct*" OR "teach*").

De nombreux tests ont été effectués pour parvenir à cette combinaison de mots-clés, afin de trouver un équilibre dans le nombre de nos résultats, entre exhaustivité et spécificité. Nous avons sélectionné cinq bases de données disponibles à travers EBSCO : APA PsycInfo, Academic Search Complete, Education Research Complete, ERIC, Psychology and Behavioral Sciences Collection. Elles ont été sélectionnées car elles regroupent des articles référencés en sciences humaines, science de l'éducation et psychologie. L'un des opérateurs d'expansion proposé par EBSCO a été utilisé pour élargir le champ de la recherche : « Appliquer des sujets équivalents ».

3.3 Critères d'éligibilité

Pour être incluses dans notre revue de littérature, les études trouvées doivent correspondre à nos trois thèmes : le flow numérique en formation. Les publications doivent être disponible en entier, en anglais ou français. Elles doivent également utilisées un instrument de mesure du flow. Ces études doivent impliquer l'usage d'un environnement numérique ou d'un outil numérique. Le dispositif étudié doit être déployé dans un contexte d'éducation ou d'apprentissage.

Aucune limite temporelle n'est précisée dans notre recherche. Les revues de littératures, dissertations et études empiriques présents dans les résultats seront exclus.

4 Résultats

Cette fonction de recherche, dont les résultats ont été enregistrés le 24 janvier 2025, a fait ressortir 427 publications (voir fig.1).

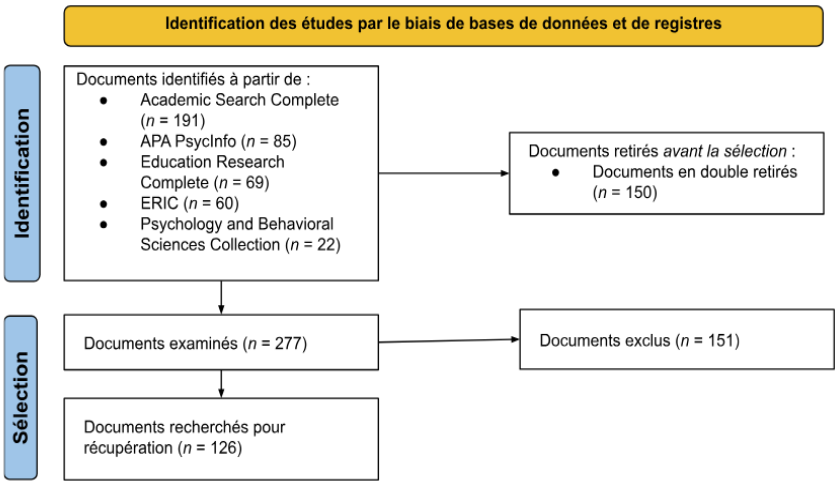


Fig. 1. Phase d'identification issue du diagramme PRISMA. D'après Page et al. (2021).

De ce corpus initial, les doublons ont été éliminés (n=150). 277 articles ont été examiné afin d'opérer une sélection basée sur les titres et les résumés : 151 articles ont été exclus, pour nous en laisser 126. Les articles restants ont été évalués selon les critères d'éligibilité définis en amont. Nous avons ainsi abouti à un corpus final à analyser. L'objet de ce poster est de partager ces premières analyses, afin de poursuivre les recherches en cours.

References

1. Agarwal, R., & Karahanna, E. (2000). Time Flies When You're Having Fun: Cognitive Absorption and Beliefs about Information Technology Usage. *MIS Quarterly*, *24*(4), 665–694.
2. Charlier, B. (2019) . Chapitre 2. Les environnements numériques d'apprentissage : quelques éléments d'intelligibilité pour la e-Formation. Dans Jézégou, A. (dir.), Traité de la e-Formation des adultes. (p. 49 -68). De Boeck Supérieur.
3. Chevry Pebayle, E., et Rossini, I. (2017) « Besoins et dispositifs d'aide dans le contexte d'apprentissage des MOOC », Distances et médiations des savoirs [En ligne], 19.
4. Choi, D. H., Kim, J., & Kim, S. H. (2007). ERP training with a web-based electronic learning system: The flow theory perspective. *International Journal of Human-Computer Studies*, *65*(3), 223–243.
5. Coleman, L. J. (2014). "Being a Teacher": emotions and optimal experience while teaching gifted children. J. Educ. Gifted 37, 56–69.

6. Csikszentmihalyi, M. (1975). Beyond boredom and anxiety. San Francisco, CA: Jossey Bass.
7. Csikszentmihalyi, M. (1990). Flow: The Psychology of Optimal Experience. New York, Harper and Row.
8. Csikszentmihalyi, M. (2014). Applications of Flow in Human Development and Education. Springer, Dordrecht : Netherlands.
9. Heutte J. (2019a). Les fondements de l'éducation positive : Perspective psychosociale et systémique de l'apprentissage. Dunod : Paris, France.
10. Heutte J. (2021). L'expérience autotélique dans les EIAH : genèse socio-historique, épistémologique et critique de l'émergence des technologies positives pour l'apprentissage. STICEF 28 (2), 27-60.
11. Ivory, J. D., and Magee, R. G. (2009). You can't take it with you? Effects of handheld portable media consoles on physiological and psychological responses to video game and movie content. Cyberpsychol. Behav. : Impact Internet Multimed. Virtual Reality Behav. Soc. 12, 291–297.
12. Molinari, G., Fenouillet, F., Lavoué, E. (2021). Éditorial du numéro spécial « Les technologies positives pour l'apprentissage ». STICEF (Sciences et Technologies de l'Information et de la Communication pour l'Éducation et la Formation), 28 (2).
13. Nakamura, J., & Csikszentmihalyi, M. (2009). Flow theory and research. In S. J. Lopez & C. R. Snyder (Eds.), Oxford handbook of positive psychology (2nd ed., pp. 195–206). Oxford University Press.
14. Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J.M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hròbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ, 372, n71.
15. Peifer, C., Wolters, G., Harmat, L., Jasmine Tan, J., et al. (2022). A Scoping Review of Flow Research, Frontiers in Psychology, section Positive Psychology, 12, 1-26.
16. Peterson, S. E., and Miller, J. A. (2004). Comparing the quality of students' experiences during cooperative learning and large-group instruction. J. Educ. Res. 97, 123–134.
17. Rathunde, K., and Csikszentmihalyi, M. (2005). Middle school Students' motivation and quality of experience: a comparison of montessori and traditional school environments. Am. J. Educ. 111, 341–371.
18. Thin, A. G., Hansen, L., and McEachen, D. (2011). Flow experience and mood states while playing body movement-controlled video games. Games Culture 6, 414–428.

La place des mécaniques ludiques dans les feedbacks : une revue systématique de la littérature

Adrien Chassaing-Monjou[1,2], Alexis Lebis[1][0000−0003−2104−8671], Mathieu Vermeulen[1], Anthony Fleury[1], and Nour El Mawas[2]

[1] IMT Nord Europe, Institut Mines-Télécom, Univ. Lille, Centre for Digital Systems, F-59000, Lille, France
{adrien.chassaing, alexis.lebis, mathieu.vermeulen, anthony.fleury}@imt-nord-europe.fr
[2] Université de Lorraine, Crem, F-57000, Metz, France
nour.el-mawas@univ-lorraine.fr

Résumé Cet article présente une revue systématique de la littérature sur la notion de feedback dans le contexte des *serious games* réalisée avec la méthode PRISMA. Celle-ci a pour objectif de faire le lien entre les feedbacks dans le contexte éducatif et les feedbacks associés au *serious game*. Ainsi, l'article propose une caractérisation de 13 mécaniques de jeu générant du feedback dans les *serious games* selon six propriétés.

Mots-clefs : Serious Games, Feedback, PRISMA, Mécaniques de jeu, Revue Systématique de la littérature.

Abstract. This article presents a systematic literature review on the concept of feedback in the context of serious games, conducted using PRISMA method. The aim is to establish a connection between feedback in the educational context and feedback associated with serious games. Therefore, the article proposes a characteristation of 13 game mechanics that generate feedback within serious games through six properties.

Keywords: Serious Games, Feedback, PRISMA, Game Mechanics, Systematic Literature Review.

1 Introduction

À l'origine défini en cybernétique par Wiener [60], le feedback a été reconnu comme étant un élément important de l'apprentissage [23]. Celui-ci a pour but d'informer un apprenant sur son état d'apprentissage et de l'aider dans ses décisions. Cette notion de feedback apparaît également dans le domaine du jeu.

Dans la littérature sur les *serious games* (SGs), objet à l'intersection du domaine ludique et éducatif, la dimension éducative a tendance à prédominer sur la dimension ludique, comme l'illustre la classification de Ratan et Ritterfeld [48] dont le seul élément ludique est la plateforme sur laquelle il se joue. Ceci se traduit également par des confusions entre la terminologie « *serious game* » et d'autres en lien également avec des applications gamifiées et les jeux éducatifs [4].

Suite à notre travail, nous décidons de définir un feedback dans les SGs comme une information destinée à un joueur-apprenant, donnée en réponse à une action ou à une absence d'action, dont le but est de l'informer de sa progression, ou de réguler son parcours, dans les dimensions ludiques et pédagogiques du SG.

Ce travail de recherche est dédié à la communauté des Environnements Informatiques pour l'Apprentissage Humain (EIAH) et aux concepteurs de jeux, aux ingénieurs pédagogiques, aux enseignants et aux chercheurs qui rencontrent des difficultés à choisir les mécaniques de jeu adaptées à leurs besoins en termes de feedback. Il a également pour but de mettre en avant le lien entre feedback pédagogique et ludique. Dans la section 2, nous mettons en avant différentes caractéristiques du feedback pédagogique. La section 3 détaille la méthode de la revue de littérature utilisée. La section 4 donne les résultats de notre revue de la littérature. Finalement, une conclusion est proposée dans la section 5.

2 État de l'art

2.1 Feedback pédagogique

Dans le contexte pédagogique, Narciss [42] propose de définir le feedback comme étant toute information obtenue en réponse à une action de l'apprenant, qui l'informe sur son état d'apprentissage ou sa progression. Narciss proposera par la suite une classification de ces feedbacks en se concentrant sur leur contenu, c'est-à-dire sur le type de connaissances présent dans ces derniers. Ce contenu se décline selon elle en deux composantes : la composante de vérification et celle d'information.

D'autres auteurs ont également proposé des classifications du feedback pédagogique, tels que Economides [15], Le [36], et Shute [52]. Contrairement à Narciss, Economides conçoit les feedbacks selon leurs buts, en particulier la vérification et l'élaboration. Ces deux points de vue ne sont pas pour autant contradictoires [36]. Shute considère le feedback comme formatif et le définit comme étant une information communiquée à l'apprenant qui a pour intention de modifier sa pensée ou son comportement dans le but d'améliorer son apprentissage.

2.2 Feedback dans les jeux sérieux

Le contexte de notre article est celui des SGs. Plusieurs définitions ont été données [12,61] et différentes classifications [29,48] ou discussions sur les différences

entre SGs, simulation et *educational game* existent [4]. Informellement, les SGs sont des jeux dont le seul but n'est pas l'amusement : ils possèdent aussi un but sérieux pouvant être l'éducation, la formation et l'apprentissage.

Dans les SGs, le feedback est considéré comme un élément de gamification essentiel. Cependant, une recherche systématique des feedbacks dans les SGs manque à la littérature [30]. Johnson et ses collaborateurs [30] proposent une classification des feedbacks dans les SGs selon le type de contenu. Leur classification reprend en particulier des éléments présents dans celle des auteurs mentionnés précédemment [42], [15], [52]. Cette classification se concentre sur l'aspect sérieux plutôt que ludique des SGs.

Différentes dimensions du feedback pédagogique peuvent être trouvées dans la littérature. Hattie et ses collaborateurs [23] considèrent le feedback personnel qui exprime une évaluation positive envers l'apprenant et le timing du feedback. Narciss [42] met également en avant le timing mais aussi la fréquence, le nombre d'essais de l'apprenant, ainsi que la modalité de présentation. Johnson et ses collaborateurs [30] se concentrent sur le type, le timing, la modalité de présentation de l'information et l'adaptation aux caractéristiques de l'apprenant.

3 Méthode PRISMA

Pour cette revue systématique de la littérature, nous avons suivi la méthode PRISMA de 2020 [44]. Les bases de données *IEEE Xplore*, *Scopus*, *ScienceDirect* et *WebOfScience* ont été utilisées. L'équation de recherche « Feedback AND (« Serious Game » OR « Serious Games ») » a été utilisée pour nos requêtes sur les abstracts, les titres et mots-clés. Nous avons visé les articles de conférences et de journaux en langue anglaise. Pour les dates de parution des articles, nous n'avons pas mis de restriction jusqu'à la date de la recherche, soit 20/12/2024. La figure 1 représente le processus de sélection des articles.

4 Résultats

Nous proposons maintenant, à travers les 33 articles retenus dans la revue systématique, six propriétés caractérisant le feedback ainsi que 13 mécaniques génératrices de ceux-ci. Plus précisément, nous avons extrait d'un côté, la manière dont une mécanique de jeu transmet au joueur-apprenant une information lui permettant d'être informé ou de se réguler, et de l'autre côté, la manière de caractériser cette mécanique.

Un tableau à double entrée (mécanique et propriété), récapitulant l'ensemble des résultats de cette revue de littérature, est disponible [5].

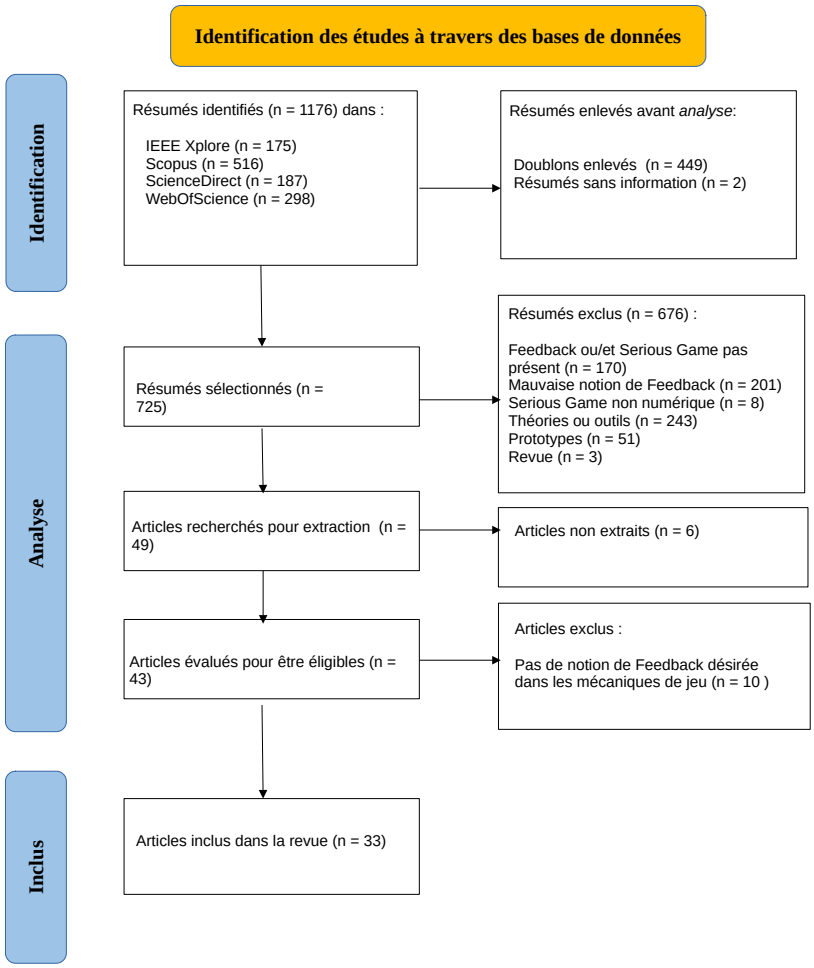


Figure 1. Diagramme de flux PRISMA pour la revue de littérature systématique des feedbacks dans les *serious games*

4.1 Propriété du Feedback des mécaniques

Nous proposons de catégoriser le feedback que ces mécaniques occasionnent selon six propriétés :

La modalité de présentation correspond à la forme avec laquelle le feedback est présenté.
La valence correspond à la valence émotionnelle de l'information contenue dans le feedback.
La temporalité correspond au timing où le feedback est donné.
La fréquence correspond à la durée d'apparition du feedback.

La compréhension correspond à la clarté sémantique du feedback.
Le but correspond au but du feedback selon notre définition donnée dans l'introduction.

4.2 Mécaniques génératrices de feedback

Lors de notre revue de la littérature, nous avons identifié 13 mécaniques de jeu, que nous avons ensuite caractérisé selon les six propriétés expliquées dans la sous-section 4.1, dans l'ordre énoncé précédemment. Pour la caractérisation suivante, nous nous nous référons à [5] :

L'assistance est, dans plus de 50% des cas, de modalité textuelle, de valence neutre, de temporalité immédiate, de fréquence ponctuelle, de compréhension explicite et de but une régulation.
La mini-map est, dans plus de 50% des cas, picturale, neutre, immédiate, continue, explicite et de régulation.
Le temps est, dans plus de 50% des cas, textuel, neutre, immédiat, continu, explicite et de régulation.
Le classement est, dans plus de 75%, représenté comme étant textuel, neutre, retardé, ponctuel, explicite et de régulation.
Le debriefing est, dans plus de 50% des cas, textuel, neutre, retardé, ponctuel, explicite et avec comme but une progression.
L'encouragement est, dans environ 50% des cas, textuel, positif, retardé, ponctuel, explicite et de régulation.
La monnaie est soit de modalité picturale ou textuelle, de valence neutre ou positive, de temporalité immédiate, de fréquence continue ou ponctuelle, de compréhension implicite ou explicite et toujours de but une progression.
Le niveau est soit textuel ou une animation et est, dans plus de 50% des cas, neutre, immédiat, continu, implicite et de but une progression.
Les points sont, dans plus de 50% des cas, textuels, négatifs, immédiats, ponctuels, explicites et de but une régulation.
La récompense est, dans environ 50% des cas, picturale, positive, retardée, ponctuelle, implicite et a comme but une progression.
Le score est, dans plus de 75% des cas, textuel, neutre, immédiat, ponctuel, explicite et de régulation.
La vie est, dans plus de 50% des cas, picturale, neutre, immédiate, continue, implicite et de régulation.

Nous voyons donc ici que des propriétés du feedback pédagogiques peuvent caractériser également des mécaniques de jeu.

5 Conclusion et Perspectives

Les résultats obtenus dans cette étude peuvent être utilisés pour aider les acteurs de la conception de jeux sérieux à choisir des mécaniques appropriées à la situation et selon les feedbacks escomptés. Notre étude montre ainsi des tendances

notables concernant les mécaniques de jeux et les propriétés des feedbacks.

Cette revue de la littérature systématique consiste en un premier travail faisant le lien entre mécaniques de jeu présentes dans les SGs et les différentes acceptions du feedback pédagogique présentes dans la littérature. À notre connaissance, ce travail reste inédit dans la communauté EIAH. Ainsi, nous avons identifié 13 mécaniques présentes dans les Gs générant du feedback. Nous avons aussi proposé six propriétés permettant de les caractériser.

Il serait pertinent par la suite de s'intéresser à d'autres éléments tels que les méthodes de conception de feedback, les méthodes d'évaluation des feedbacks ou encore les outils techniques pour générer les feedbacks. Également, il est possible de trouver dans la littérature des évaluations des différents feedbacks pédagogiques décrits par nos six propriétés, il serait intéressant de faire le lien entre ces études et les feedbacks présents dans les mécaniques de jeu.

Remerciements. Cette contribution scientifique a été réalisée dans le cadre du projet de recherche ANR-23-CE38-0001 TALE4GDA et ANR-22-CE38-0003 COPCOT.

Références

1. Altomari, L., Altomari, N., Iazzolino, G. : Gamification and Soft Skills Assessment in the Development of a Serious Game : Design and Feasibility Pilot Study. JMIR Serious Games **11** (Jul 2023). https://doi.org/10.2196/45436, https://games.jmir.org/2023/1/e45436
2. Binder, J.C., Zöllig, J., Eschen, A., Mérillat, S., Röcke, C., Schoch, S.F., Jäncke, L., Martin, M. : Multi-domain training in healthy old age : Hotel Plastisse as an iPad-based serious game to systematically compare multi-domain and single-domain training. Frontiers in Aging Neuroscience **7** (Jul 2015). https://doi.org/10.3389/fnagi.2015.00137, http://journal.frontiersin.org/Article/10.3389/fnagi.2015.00137/abstract
3. Cano, S., Naranjo, J.S., Henao, C., Rusu, C., Albiol-Pérez, S. : Serious Game as Support for the Development of Computational Thinking for Children with Hearing Impairment. Applied Sciences **11**(1), 115 (Dec 2020). https://doi.org/10.3390/app11010115, https://www.mdpi.com/2076-3417/11/1/115
4. Carrión, M., Santórum, M., Flores, H., Aguilar, J., Pérez, M. : Serious game, gamified applications, educational software : A comparative study. 2019 International Conference on Information Systems and Software Technologies (ICI2ST) pp. 55–62 (2019)
5. CHASSAING-MONJOU, A. : Jeu de données de la synthèse des mécaniques de jeux sérieux (2025). https://doi.org/10.57745/KWMUUB, https://doi.org/10.57745/KWMUUB
6. Chatzitofis, A., Monaghan, D., Mitchell, E., Honohan, F., Zarpalas, D., O'Connor, N.E., Daras, P. : HeartHealth : A Cardiovascular Disease Home-based Rehabilitation System. Procedia Computer Science **63**, 340–347 (2015). https://doi.org/10.1016/j.procs.2015.08.352, https://linkinghub.elsevier.com/retrieve/pii/S1877050915024874
7. Conde, S., Elor, A., Kurniawan, S. : Boundaries : A Serious Game on Relationships for Individuals with Developmental Disabilities. In : 2020 IEEE 8th International Conference on Serious Games and Applications for Health (SeGAH). pp. 1–7. IEEE, Vancouver, BC, Canada (Aug 2020). https://doi.org/10.1109/SeGAH49190.2020.9201810, https://ieeexplore.ieee.org/document/9201810/
8. Covaci, A., Ghinea, G., Lin, C.H., Huang, S.H., Shih, J.L. : Multisensory games-based learning - lessons learnt from olfactory enhancement of a digital board game. Multimedia Tools and Applications **77**(16) (Aug 2018). https://doi.org/10.1007/s11042-017-5459-2, http://link.springer.com/10.1007/s11042-017-5459-2
9. Dell'Aquila, E., Vallone, F., Zurlo, M.C., Marocco, D. : SG-ACCORD : Designing Virtual Agents for Soft Skills Training in the School Context. Education Sciences **12**(3), 174 (Mar 2022). https://doi.org/10.3390/educsci12030174, https://www.mdpi.com/2227-7102/12/3/174
10. Dietvorst, E., Aukes, M.A., Legerstee, J.S., Vreeker, A., Hrehovcsik, M.M., Keijsers, L., Hillegers, M.H.J. : A Smartphone Serious Game for Adolescents (Grow It ! App) : Development, Feasibility, and Acceptance Study. JMIR Formative Research **6**(3) (Mar 2022). https://doi.org/10.2196/29832, https://formative.jmir.org/2022/3/e29832
11. Djaouti, D., Alvarez, J., Jessel, J.P. : Classifying serious games : The g/p/s model. In : Handbook of Research on Improving Learning and Motivation through Educational Games (2011)

12. Djaouti, D., Alvarez, J., Jessel, J.P., Rampnoux, O. : Origins of serious games. In : Serious Games and Edutainment Applications (2011)
13. Dorrestein, L., Jansen, J., Plagis, T., Ritter, C., Vertenten, G., Barkema, H.W. : Use of an online gaming tool, the Veterinary DialogueTrainer, for teaching clinical communication skills to bovine veterinary practitioners. Frontiers in Veterinary Science **10** (Jul 2023). https://doi.org/10.3389/fvets.2023.1192598, https://www.frontiersin.org/articles/10.3389/fvets.2023.1192598/full
14. Duarte, N., Postolache, O., Scharcanski, J. : KSGphysio - Kinect serious game for physiotherapy. In : 2014 International Conference and Exposition on Electrical and Power Engineering (EPE). pp. 606–611. IEEE, Iasi, Romania (Oct 2014). https://doi.org/10.1109/ICEPE.2014.6969981, http://ieeexplore.ieee.org/document/6969981/
15. Economides, A.A. : Adaptive feedback evaluation. In : Proceedings of the 5th WSEAS International Conference on Distance Learning and Web Engineering (2005)
16. Er, C.C.W., Lau, B.T., Zheng, P. : An Audio and Haptic Feedback-Based Virtual Environment Spatial Navigation Learning Tool. In : 2018 Joint 10th International Conference on Soft Computing and Intelligent Systems (SCIS) and 19th International Symposium on Advanced Intelligent Systems (ISIS). pp. 741–746. IEEE, Toyama, Japan (Dec 2018). https://doi.org/10.1109/SCIS-ISIS.2018.00123, https://ieeexplore.ieee.org/document/8716235/
17. Estrella, T., Alfonso, C., Ramos-Castro, J., Alsina, A., Capdevila, L. : A Serious Game to Self-Regulate Heart Rate Variability as a Technique to Manage Arousal Level Through Cardiorespiratory Biofeedback : Development and Pilot Evaluation Study. JMIR Serious Games **11** (Aug 2023). https://doi.org/10.2196/46351, https://games.jmir.org/2023/1/e46351
18. Franzwa, C., Tang, Y., Johnson, A., Bielefeldt, T. : Balancing Fun and Learning in a Serious Game Design :. International Journal of Game-Based Learning **4**(4), 37–57 (Oct 2014). https://doi.org/10.4018/ijgbl.2014100103, https://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/ijgbl.2014100103
19. de Freitas, S., Liarokapis, F. : Serious games : A new paradigm for education ? In : Serious Games and Edutainment Applications (2011)
20. Ganzeboom, M., Bakker, M., Beijer, L., Rietveld, T., Strik, H. : Speech training for neurological patients using a serious game. British Journal of Educational Technology **49**(4), 761–774 (Jul 2018). https://doi.org/10.1111/bjet.12640, https://bera-journals.onlinelibrary.wiley.com/doi/10.1111/bjet.12640
21. Ghodousi Moghadam, S., Mazloum Khorasani, Z., Sharifzadeh, N., Tabesh, H. : A mobile serious game about diabetes self-management : Design and evaluation. Heliyon **10**(18) (Sep 2024). https://doi.org/10.1016/j.heliyon.2024.e37755, https://linkinghub.elsevier.com/retrieve/pii/S2405844024137863
22. Ghodsvali, M., Dane, G., De Vries, B. : An online serious game for decision-making on food-water-energy nexus policy. Sustainable Cities and Society **87** (Dec 2022). https://doi.org/10.1016/j.scs.2022.104220, https://linkinghub.elsevier.com/retrieve/pii/S221067072200525X
23. Hattie, J.A.C., Timperley, H.S. : The power of feedback. Review of Educational Research **77**, 112 – 81 (2007)
24. Healy, P., You, K.L., Xu, A., Thomas, T.H., Babichenko, D. : A Narrative Serious Game to Teach Self-Advocacy Skills in Advanced Cancer. Procedia Computer Science **206**, 162–172 (2022). https://doi.org/10.1016/j.procs.2022.09.095, https://linkinghub.elsevier.com/retrieve/pii/S1877050922009681

25. Heins, S., Dehem, S., Montedoro, V., Dehez, B., Edwards, M., Stoquart, G., Rocca, F., De Deken, P., Mancas, M., Lejeune, T. : Robotic-assisted serious game for motor and cognitive post-stroke rehabilitation. In : 2017 IEEE 5th International Conference on Serious Games and Applications for Health (SeGAH). pp. 1–8. IEEE, Perth, WA (Apr 2017). https://doi.org/10.1109/SeGAH.2017.7939262, http://ieeexplore.ieee.org/document/7939262/
26. Hoelbling, D., Salmhofer, A., Gencoglu, C., Baranyi, R., Pinter, K., Özbay, S., Ulupinar, S., Ozkara, A.B., Grechenig, T. : JudgED : Comparison between Kickboxing Referee Performance at a Novel Serious Game for Judging Improvement and at World Championships. Applied Sciences **13**(17) (Aug 2023). https://doi.org/10.3390/app13179549, https://www.mdpi.com/2076-3417/13/17/9549
27. Hu, L., Zhang, L., Yin, R., Li, Z., Shen, J., Tan, H., Wu, J., Zhou, W. : NEOGAMES : A Serious Computer Game That Improves Long-Term Knowledge Retention of Neonatal Resuscitation in Undergraduate Medical Students. Frontiers in Pediatrics **9** (Apr 2021). https://doi.org/10.3389/fped.2021.645776, https://www.frontiersin.org/articles/10.3389/fped.2021.645776/full
28. Ishac, K., Bourahmoune, K., Carmichael, M. : An IoT Sensing Platform and Serious Game for Remote Martial Arts Training. Sensors **23**(17) (Aug 2023). https://doi.org/10.3390/s23177565, https://www.mdpi.com/1424-8220/23/17/7565
29. Jantke, K.P., Gaudl, S.E. : Taxonomic contributions to digital games science. 2010 2nd International IEEE Consumer Electronics Society's Games Innovations Conference pp. 1–8 (2010)
30. Johnson, C.I., Bailey, S.K.T., Buskirk, W.L.V. : Designing effective feedback messages in serious games and simulations : A research review. In : Instructional Techniques to Facilitate Learning and Motivation of Serious Games. pp. 119–140 (2017)
31. Jouan, J., De Graeuwe, M., Carof, M., Baccar, R., Bareille, N., Bastian, S., Brogna, D., Burgio, G., Couvreur, S., Cupiał, M., Dumont, B., Jacquot, A.L., Magagnoli, S., Makulska, J., Maréchal, K., Pérès, G., Ridier, A., Salou, T., Tombarkiewicz, B., Sgolastra, F., Godinot, O. : Learning Interdisciplinarity and Systems Approaches in Agroecology : Experience with the Serious Game SEGAE. Sustainability **12**(11) (May 2020). https://doi.org/10.3390/su12114351, https://www.mdpi.com/2071-1050/12/11/4351
32. Jurivich, D.A., Eck, R.V., Wood, J.R., Snustad, D., Holloway, J., Langendoen, D., Hughes, M., Galynker, B., Brewster, S., Manocha, G.D. : A geriatric game to strengthen older adult healthcare delivery and outcomes. Journal of the American Geriatrics Society (2024), https://api.semanticscholar.org/CorpusID:269633922
33. Kapp, F., Spangenberger, P., Kruse, L., Narciss, S. : Investigating changes in self-evaluation of technical competences in the serious game Serena Supergreen : Findings, challenges and lessons learned. Metacognition and Learning **14**(3), 387–411 (Dec 2019). https://doi.org/10.1007/s11409-019-09209-4, http://link.springer.com/10.1007/s11409-019-09209-4
34. Knobel, S.E.J., Kaufmann, B.C., Gerber, S.M., Urwyler, P., Cazzoli, D., Müri, R.M., Nef, T., Nyffeler, T. : Development of a Search Task Using Immersive Virtual Reality : Proof-of-Concept Study. JMIR Serious Games **9**(3) (Jul 2021). https://doi.org/10.2196/29182, https://games.jmir.org/2021/3/e29182
35. Laamarti, F., Eid, M.A., Saddik, A.E. : An overview of serious games. Int. J. Comput. Games Technol. **2014** (2014)
36. Le, N.T. : A classification of adaptive feedback in educational systems for programming. Syst. **4**, 22 (2016)

37. Lee, A., Goodman, S., Chen, C.M., Landau, R., Chatterji, M. : Electronic Feedback Alone Versus Electronic Feedback Plus in-Person Debriefing for a Serious Game Designed to Teach Novice Anesthesiology Residents to Perform General Anesthesia for Cesarean Delivery : Randomized Controlled Trial. JMIR Serious Games **12** (Nov 2024). https://doi.org/10.2196/59047, https://games.jmir.org/2024/1/e59047
38. Mohan, D., Rosengart, M.R., Fischhoff, B., Angus, D.C., Farris, C., Yealy, D.M., Wallace, D.J., Barnato, A.E. : Testing a videogame intervention to recalibrate physician heuristics in trauma triage : study protocol for a randomized controlled trial. BMC Emergency Medicine **16**(1), 44 (Dec 2016). https://doi.org/10.1186/s12873-016-0108-z, https://bmcemergmed.biomedcentral.com/articles/10.1186/s12873-016-0108-z
39. Mueller, S., Soriano, D., Boscor, A., Saville, N., Arjyal, A., Baral, S., Fordham, M., Hearn, G., Le Masson, V., Kayastha, R., Kostkova, P. : MANTRA : development and localization of a mobile educational health game targeting low literacy players in low and middle income countries. BMC Public Health **20**(1) (Dec 2020). https://doi.org/10.1186/s12889-020-09246-8, https://bmcpublichealth.biomedcentral.com/articles/10.1186/s12889-020-09246-8
40. Mullinix, G., Gray, O., Colado, J., Veinott, E., Leonard, J., Papautsky, E.L., Argenta, C., Clover, M., Sickles, S., Castronova, E., Todd, P.M., Ross, T., Lorince, J., Hoteling, J., Mayell, S., Hale, C., Whitaker, E., Hoffman, R.R., Fox, O., Flach, J. : Heuristica : Designing a serious game for improving decision making. In : 2013 IEEE International Games Innovation Conference (IGIC). pp. 250–255. IEEE, Vancouver, BC, Canada (Sep 2013). https://doi.org/10.1109/IGIC.2013.6659159, http://ieeexplore.ieee.org/document/6659159/
41. Muratet, M., Torguet, P., Viallet, F., Jessel, J. : Experimental Feedback on Prog&Play : A Serious Game for Programming Practice. Computer Graphics Forum **30**(1), 61–73 (Mar 2011). https://doi.org/10.1111/j.1467-8659.2010.01829.x, https://onlinelibrary.wiley.com/doi/10.1111/j.1467-8659.2010.01829.x
42. Narciss, S. : Feedback strategies for interactive learning tasks. In : Handbook of Research on Educational Communications and Technology. pp. 125–144 (2007)
43. Nasirzade, A., Deldar, K., Froutan, R., Shakeri, M.T. : Comparison of the effects of burn assessment mission game with feedback lecture on nursing students' knowledge and skills in the burn patients' assessment : a randomized clinical trial. BMC Medical Informatics and Decision Making **24**(1), 157 (Jun 2024). https://doi.org/10.1186/s12911-024-02558-4, https://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-024-02558-4
44. Page, M.J., McKenzie, J.E., Bossuyt, P.M.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., Shamseer, L., Tetzlaff, J.M., Akl, E.A., Brennan, S.E., Chou, R., Glanville, J.M., Grimshaw, J.M., Hrõbjartsson, A., Lalu, M.M., Li, T., Loder, E.W., Mayo-Wilson, E., McDonald, S., McGuinness, L.A., Stewart, L.A., Thomas, J., Tricco, A.C., Welch, V.A., Whiting, P.F., Moher, D. : The prisma 2020 statement : an updated guideline for reporting systematic reviews. Systematic Reviews **10** (2020), https://api.semanticscholar.org/CorpusID:232387120
45. Paquette, G. : Feedback, rétroaction, rétroinformation, réponse... du pareil au même. Communication Et Langages **73**, 5–18 (1987), https://api.semanticscholar.org/CorpusID:144733351

46. Paracha, S., Jehanzeb, S., Yoshie, O. : A Serious Game for Inculcating Islamic Values in Children. In : 2013 Taibah University International Conference on Advances in Information Technology for the Holy Quran and Its Sciences. pp. 172–177. IEEE, Madinah, Saudi Arabia (Dec 2013). https://doi.org/10.1109/NOORIC.2013.44, http://ieeexplore.ieee.org/document/7277241/
47. Pasquier, P., Mérat, S., Malgras, B., Petit, L., Queran, X., Bay, C., Boutonnet, M., Jault, P., Ausset, S., Auroy, Y., Perez, J.P., Tesnière, A., Pons, F., Mignon, A. : A Serious Game for Massive Training and Assessment of French Soldiers Involved in Forward Combat Casualty Care (3D-SC1) : Development and Deployment. JMIR Serious Games **4**(1) (May 2016). https://doi.org/10.2196/games.5340, http://games.jmir.org/2016/1/e5/
48. Ratan, R.A., Ritterfeld, U. : Classifying serious games. In : Serious Games (2009)
49. Sarasmita, M.A., Larasanty, L.P.F., Kuo, L.N., Cheng, K.J., Chen, H.Y. : A Computer-Based Interactive Narrative and a Serious Game for Children With Asthma : Development and Content Validity Analysis. Journal of Medical Internet Research **23**(9) (Sep 2021). https://doi.org/10.2196/28796, https://www.jmir.org/2021/9/e28796
50. Schättin, A., Pickles, J., Flagmeier, D., Schärer, B., Riederer, Y., Niedecken, S., Villiger, S., Jurt, R., Kind, N., Scott, S.N., Stettler, C., Martin-Niedecken, A.L. : Development of a Novel Home-Based Exergame With On-Body Feedback : Usability Study. JMIR Serious Games **10**(4) (Dec 2022). https://doi.org/10.2196/38703, https://games.jmir.org/2022/4/e38703
51. Sharif, K.H., Ameen, S.Y. : A Intelligent Security Power Lab (SPL) : The Ultimate Serious Game Training in Cybersecurity pp. 245–259 (2023)
52. Shute, V.J. : Focus on formative feedback. Review of Educational Research **78**, 153 – 189 (2007)
53. Sriwatanathamma, P., Sirivesmas, V., Simatrang, S., Bhowmik, N.H. : Gamifying Cognitive Behavioral Therapy Techniques on Smartphones for Bangkok's Millennials With Depressive Symptoms : Interdisciplinary Game Development. JMIR Serious Games **11** (May 2023). https://doi.org/10.2196/41638, https://games.jmir.org/2023/1/e41638
54. Stach, C., Schlindwein, L.F.M. : Candy Castle &#x2014 ; A prototype for pervasive health games. In : 2012 IEEE International Conference on Pervasive Computing and Communications Workshops. pp. 501–503. IEEE, Lugano, Switzerland (Mar 2012). https://doi.org/10.1109/PerComW.2012.6197547, http://ieeexplore.ieee.org/document/6197547/
55. Starkov, M., Jochems, S., Rijsdijk, J., Snellenberg, R., Stoffels, L., Zaidi, A., Bidarra, R. : Sonifying motor skills with Pizzicato, a game for motor behavior research. In : 2024 IEEE Conference on Games (CoG). pp. 1–8. IEEE, Milan, Italy (Aug 2024). https://doi.org/10.1109/CoG60054.2024.10645594, https://ieeexplore.ieee.org/document/10645594/
56. Su, J.M., Yang, Y.C., Weng, T.N., Li, M.J., Wang, C.J. : A web-based serious game about self-protection for COVID-19 prevention : Development and usability testing. Comunicar **29**(69), 91–111 (Oct 2021). https://doi.org/10.3916/C69-2021-08, https://www.revistacomunicar.com/index.php?contenido=detalles&numero=69&articulo=69-2021-08
57. Suppan, M., Catho, G., Robalo Nunes, T., Sauvan, V., Perez, M., Graf, C., Pittet, D., Harbarth, S., Abbas, M., Suppan, L. : A Serious Game Designed to Promote Safe Behaviors Among Health Care Workers During the COVID-19 Pandemic : Development of "Escape COVID-19". JMIR Serious Games **8**(4) (Dec 2020). https://doi.org/10.2196/24986, http://games.jmir.org/2020/4/e24986/

58. Szczurowski, K., Smith, M. : "Woodlands" - a Virtual Reality Serious Game Supporting Learning of Practical Road Safety Skills. In : 2018 IEEE Games, Entertainment, Media Conference (GEM). pp. 1–9. IEEE, Galway (Aug 2018). https://doi.org/10.1109/GEM.2018.8516493, https://ieeexplore.ieee.org/document/8516493/
59. Volioti, C., Tsagaris, A., Trigkas, D., Iliou, T., Katsantonis, M.N., Mavridis, I. : HapticSOUND : An Interactive Learning Experience with a Digital Musical Instrument. Applied Sciences **13**(12) (Jun 2023). https://doi.org/10.3390/app13127149, https://www.mdpi.com/2076-3417/13/12/7149
60. Wiener, N. : Cybernetics or control and communication in the animal and the machine. MIT Press (1948)
61. Zyda, M.J. : From visual simulation to virtual reality to games. Computer **38**, 25–32 (2005)

Méthode de comparaison des modèles de jeux sérieux : vers un recouvrement sémantique des propositions de la littérature

Madeleine Valat, Alexis Lebis et Anthony Fleury

IMT Nord Europe, Institut Mines-Télécom, Univ. Lille, Centre for Digital Systems
F-59000 Lille, France
prenom.nom@imt-nord-europe.fr

Résumé. Suscitant l'intérêt des élèves et facilitant l'apprentissage, les jeux sérieux sont des outils de choix pour l'enseignement. Ainsi, de nombreux modèles ont été proposés, qui mettent en avant leurs dimensions pédagogique et ludique. Il n'existe cependant pas de consensus sur les taxonomies utilisées dans ces modèles, ce qui rend difficile leur comparaison. Nous proposons donc une manière de classifier les concepts de ces modèles via une approche sémantique, qui sera ici centrée sur la pédagogie.

Mots-clés : Jeu sérieux , Pédagogie , Sémantique , Comparaison , Méthode

Abstract. By stimulating student interest and facilitating learning, serious games are ideal teaching tools. Numerous models have been proposed, highlighting their pedagogical and ludic dimensions. However, there is no consensus on the taxonomies used in these models, which makes it difficult to compare them. We therefore propose a way of classifying the concepts of these models via a semantic approach, which will be focused here on pedagogy.

Keywords: Serious game , Pedagogy , Semantics , Comparison , Method

1 Introduction et contexte

Face aux changements d'attentes des générations par rapport à l'apprentissage, les jeux sérieux (JS) sont aujourd'hui en plein essor. Il est admis par la communauté que ceux-ci combinent à la fois des éléments ludiques et des éléments pédagogiques, ces deux aspects étant étudiés parallèlement ou séparément dans la littérature existante.

Les travaux présentés ici visent à étudier la représentativité des aspects pédagogique et ludique dans les modèles de JS (MJS), afin de pouvoir comparer ceux-ci sémantiquement. Si plusieurs auteurs ont abordé ces deux dimensions, nous remarquons une divergence dans les termes utilisés et la couverture de ces modèles. Il ne semble pas y avoir de consensus au niveau des taxonomies utilisées, ce qui interroge sur les propriétés des différents modèles proposés, ainsi qu'à leur contexte d'utilisation et conséquences associées.

Devant ce constat, nous posons deux questions de recherche :

1. Comment comparer les différents MJS ?
2. La couverture des MJS est-elle équivalente ?

Dans ce papier, nous étudierons ces deux questions via la dimension pédagogique. Pour cela, après un état de l'art (section 2), nous décrirons la méthode utilisée (section 3) avant d'exposer nos résultats (section 4) et de conclure.

2 État de l'art

Dans des revues ou études de la littérature, Slimani et al. [15], Tori et al. [16] et Ávila-Pesántez et al. [3] ont comparé des MJS selon différentes dimensions. Cependant, ces travaux se situant à un niveau macroscopique, les similarités et disparités inter-modèles du vocabulaire ne sont pas étudiées.

Ici, nous comparons les MJS présentant un aspect ludique et un aspect pédagogique, qu'ils soient orientés conception [11,12,6], classification [10,7,8], ou vers une compréhension plus générale des JS [1,4,3,2,13,14]. Ces douze MJS font ressortir différents aspects des JS. Ainsi, [11,12,7,8,1,4,14,13,3] s'intéressent au contexte d'utilisation du JS selon différentes approches (matériel, utilisation, domaine, expérience, informations techniques, ...). [11,12,6,10,4,2] présentent une notion d'ordonnancement des activités ludiques, et [11,7,8,14,13,2] considèrent la manière de jouer, sous des acceptions hétérogènes (ressorts de jeu, mécaniques, gameplay, ...). [7,13,12,4,3,6] mentionnent également un objectif pédagogique, et [11,10,14,2,1,4,3,6] utilisent le concept d'activité pédagogique via différentes approches (fonction, type, compétences, théorie, ...).

Ainsi, nous constatons une grande hétérogénéité dans les MJS. La littérature sur la comparaison de modèles [17] suggère d'utiliser conjointement une approche structurelle et une approche sémantique. Ici, nous chercherons donc à comparer les modèles sémantiquement (i.e. à confronter leurs vocabulaires). L'approche structurelle fera l'objet d'une publication à part.

3 Méthode de comparaison sémantique

3.1 Utilisation d'une taxonomie hiérarchisée

L'un des verrous identifiés dans la comparaison sémantique des MJS est la diversité des termes utilisés. Devant cela, nous avons décidé de les étudier en utilisant une taxonomie organisée en thésaurus, nous permettant une catégorisation explicite. Nous utilisons ici le Thésaurus Européen d'Éducation (TEE) [9] et éduthès [5] comme pivots de comparaison, thésauri produits par des autorités faisant référence.

Les thésauri étant des dictionnaires hiérarchisés, nous utilisons cette propriété pour définir des *catégories* (cf Figure 1), et classifier les *termes*.

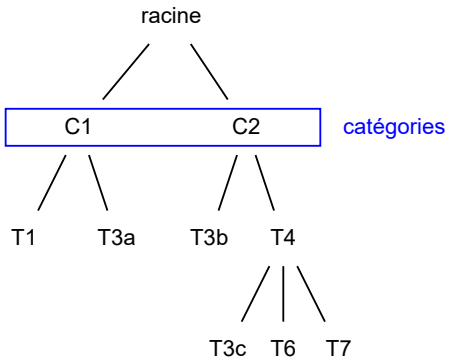


FIG. 1 – Structuration d'un thésaurus et définition des catégories

3.2 Métriques de comparaison

Pour chaque terme considéré (TC), nous lui associons une liste pondérée de catégories, dépendante des correspondances trouvées dans le thésaurus pour ce terme. Nous distinguons trois types de correspondance, sans considération pour leur profondeur dans le thésaurus :

Exacte Le TC est présent, le poids associé à la catégorie est de 1.

Partielle Le TC exact est inexistant mais des termes approchés (i.e. des termes renvoyés comme résultats à la recherche de TC) sont identifiés. Le poids associé à une catégorie sera alors de $\frac{a}{n+1}$ (avec n le nombre total de termes approchés et a le nombre de ces termes appartenant à la catégorie considérée).

Absence Aucune catégorie n'est associée à TC.

TABLEAU 1 – Exemple de catégorisation des termes T1, T2 et T3 selon le thésaurus présenté Figure 1

Terme	Catégorie	Poids
T1	C1	1.0
T2	$\emptyset$	$\emptyset$
T3	C1	0.25 (a=1,n=3)
	C2	0.5 (a=2, n=3)

Après la répartition des termes considérés dans les catégories, nous analysons la fréquence d'apparition de celles-ci dans chaque modèle. Nous définissons alors les coefficients $c^{+}_{M,C}$ et $c^{\times}_{M,C}$ correspondant au modèle M et à la catégorie C

par :

$$c^{+}_{M,C} = \frac{\sum_{t \in M} c_{t,C}}{\text{card}(M)} \tag{1}$$

$$c^{\times}_{M,C} = \prod_{t \in M} c_{t,C} \tag{2}$$

où $c_{t,C}$ est la valeur associée au terme t et à la catégorie C.

Ces coefficients permettent d'avoir deux méthodes de comparer la représentativité d'un modèle par rapport à une catégorie : une compensatoire (additive), pour connaître l'expressivité d'une catégorie pour un modèle (Equation 1), et une semi-compensatoire (multiplicative), reflétant l'ambiguïté sémantique d'un modèle dans la taxonomie (Equation 2).

4 Résultats

Les résultats présentés ici comparent sémantiquement à éduthès [5] les MJS de [1,2,3,4,6,7,8,10,13,11,12,14]. [1] Les quartiles (Q1, Q3), la médiane (Q2), le maximum et l'intervalle théorique des valeurs sont indiqués sous les tableaux.

4.1 Expressivité des modèles

En approche compensatoire, nous obtenons pour [7] une valeur élevée en "Gestion de l'information" et une faible en "Apprentissage/Cognition" (Table 2). Ceci met en évidence la focalisation de ce modèle sur la description de l'aspect éducationnel plutôt que sur la manière utilisée pour cela. Nous interprétons la faible valeur de [8] en "Approches et méthodes d'enseignement" comme une non-considération des méthodes utilisées par l'enseignant.

Ces données semblent démontrer la capacité de notre méthode à décrire formellement les MJS, ce qui serait un résultat significatif dans nos recherches.

Tableau 2 – Expressivité des modèles (extrait)

Catégorie	[7]	[13]	[8]
Apprentissage/Cognition	0.068	0.873	0.517
Approches et méthodes d'enseignement	0.427	0.476	0.017
Gestion de l'information	0.980	0.119	0.026
Gestion scolaire	0.175	0.655	0.085
Intervalle=[0; 10] Q1=0.04275	Q2=0.174	Q3=0.41175	Q4=1.535

1. L'intégralité des résultats est sur https://doi.org/10.57745/GJDVV9 [18]. Les quatre tableaux présentent les coefficients obtenus avec les données du TEE [9] et d'éduthés [5], d'après les méthodes compensatoire et semi-compensatoire.

4.2 Ambiguïté sémantique

En semi-compensatoire (Table 3), nous remarquons que [7] est associé à beaucoup de valeurs très faibles, qu'il est très ambigu sémantiquement. Cela peut signifier qu'il est peu pertinent de décrire ce modèle selon ce thésaurus, ou du moins de décrire certains termes qui seraient associés à de très petites valeurs. A contrario, [8] et [13] sont sémantiquement clairs, ce qui confirme la pertinence de leur expressivité sémantique.

Tableau 3 – Ambiguïté des modèles (extrait)

Catégorie	[7]	[13]	[8]
Apprentissage/Cognition	10^{-3}	0.167	0.034
Approches et méthodes d'enseignement	10^{-3}	1.000	0.034
Gestion de l'information	0.007	0.250	0.051
Gestion scolaire	10^{-3}	0.034	0.171
Intervalle=[0; 1] Q1=0.00525	Q2=0.0465	Q3=0.192	Q4=1.000

4.3 Vue d'ensemble

Nous remarquons dans la Table 4 que des valeurs d'expressivité élevées sont associées à des valeurs faibles pour l'ambiguïté. Nous proposons donc d'analyser l'ambiguïté sémantique en comparant des catégories ayant des valeurs d'expressivité similaires (comme "Éducation (Général)" et "Apprentissage/Cognition").

Ce tableau montre une forte représentativité des catégories "Milieu", "Éducation (général)", "Évaluation", ainsi que "Apprentissage/Cognition" et "Approches et méthodes d'enseignement" dans les différents MJS. La présence des deux dernières catégories tend à démontrer l'importance de la pédagogie dans les MJS.

5 Conclusion et perspectives

Dans ce papier, nous proposons une méthode de comparaison des MJS selon leur dimension éducationnelle via les notions d'expressivité et d'ambiguïté sémantique. Les résultats obtenus nous permettent de décrire les modèles selon différentes catégories de l'éducation, avec une analyse conjointe des deux dimensions sur les termes. Si les données obtenues ont révélé des consensus, il serait pertinent d'affiner notre métrique pour les correspondances partielles (afin d'éviter les très petites valeurs d'ambiguïté sémantique), de modifier le calcul du coefficient compensatoire (pour que les données soient mieux réparties dans l'intervalle), ainsi que de trouver une manière de tenir compte des termes composés de plusieurs mots. Il serait également intéressant de comparer les résultats obtenus sur différents thésauri.

Tableau 4 – Représentativité des catégories

Catégorie	Expressivité $(c^{+}_{M,C})$	Ambiguïté $(c^{\times}_{M,C})$	Modèles concernés
Administration/Relations de travail	0.977	$10^{-10.5}$	7
Apprentissage/Cognition	4.544	$10^{-17.9}$	9
Approches et méthodes d'enseignement	3.125	$10^{-20.8}$	10
Caractéristiques psychologiques et physiques	4.236	$10^{-18.1}$	10
Entente intergouvernementale	0.006	$10^{-1.7}$	1
Gestion de l'information	1.572	$10^{-7.2}$	6
Gestion scolaire	2.021	$10^{-15.9}$	10
Identificateur	0	0	0
Individus	0.554	$10^{-7.8}$	4
Matériel didactique	1.383	$10^{-9.4}$	7
Milieu	4.590	$10^{-15.6}$	7
Orientation scolaire et professionnelle	0.047	$10^{-4.4}$	3
Programmes d'études/Disciplines	3.338	$10^{-15.9}$	9
Recherche et développement	0.429	$10^{-3.9}$	4
Services connexes	0.803	$10^{-11.1}$	7
Types de documents	0.979	$10^{-16.3}$	10
Éducation (général)	4.128	$10^{-22.1}$	10
Établissements d'enseignement	0.614	$10^{-7.1}$	7
Évaluation	4.411	$10^{-21.4}$	11
Intervalle	$[0; 120]$	$[0; 1]$	

Nos travaux ultérieurs se concentreront principalement sur l'extension de cette comparaison aux aspects ludiques qui, combinée à une approche structurelle, nous permettra de confronter les MJS globalement.

Remerciement

Ce travail est financé par le projet ANR TALE4GDA (ANR-23-CE38-0001).

Références

1. Amory, A. : Game object model version II : A theoretical framework for educational game development. Educational Technology Research and Development **55**(1), 51–77 (Feb 2007). https://doi.org/10.1007/s11423-006-9001-x
2. Arnab, S., Lim, T., Carvalho, M.B., Bellotti, F., de Freitas, S., Louchart, S., Suttie, N., Berta, R., De Gloria, A. : Mapping learning and game mechanics for serious games analysis. British Journal of Educational Technology **46**(2), 391–411 (2015). https://doi.org/10.1111/bjet.12113
3. Ávila-Pesántez, D., Rivera, L.A., Alban, M.S. : Approaches for Serious Game Design : A Systematic Literature Review **8**(3) (2017)

4. Carvalho, M.B., Bellotti, F., Berta, R., Gloria, A.D., Sedano, C.I., Hauge, J.B., Hu, J., Rauterberg, M. : An activity theory-based model for serious games analysis and conceptual design. Computers & Education **87**, 166–181 (2015). https://doi.org/10.1016/j.compedu.2015.03.023
5. de documentation collégiale (CDC), C. : EDUthès : Thésaurus de l'éducation. https ://eduthes.cdc.qc.ca/vocab/ (Nov 2016), last accessed : 2025-01-28
6. De Troyer, O., Van Broeckhoven, F., Vlieghe, J. : Linking serious game narratives with pedagogical theories and pedagogical design strategies. Journal of Computing in Higher Education **29**, 549–573 (2017)
7. Djaouti, D., Alvarez, J., Jessel, J.P. : Classifying serious games : The G/P/S model. In : Handbook of Research on Improving Learning and Motivation through Educational Games : Multidisciplinary Approaches, pp. 118–136. Advances in Game-Based Learning, IGI global (2011)
8. Elborji, Y., Khaldi, M. : An IEEE LOM Application Profile to Describe Serious Games «SG-LOM». International Journal of Computer Applications **86**(13), 1–8 (Jan 2014). https://doi.org/10.5120/15042-3404
9. of Europe, C. : European Education Thesaurus. https ://vocabularyserver.com/tee/en/ (Dec 2003), last accessed : 2025-01-28
10. Ildephonce, I., Mugisa, E., Allen, C. : Learning Objects in Instructional Serious Game Design. In : 2018 IEEE 18th International Conference on Advanced Learning Technologies (ICALT). pp. 119–121. IEEE Computer Society (Jul 2018). https://doi.org/10.1109/ICALT.2018.00035
11. Mariais, C. : Modèles pour la conception de Learning Role-Playing Games en formation professionnelle. Ph.D. thesis, Université de Grenoble (Apr 2012)
12. Marne, B., Wisdom, J., Huynh-Kim-Bang, B., Labat, J.M. : The Six Facets of Serious Game Design : A Methodology Enhanced by Our Design Pattern Library. In : Ravenscroft, A., Lindstaedt, S., Kloos, C.D., Hernández-Leo, D. (eds.) 21st Century Learning for 21st Century Skills. pp. 208–221. Springer Berlin Heidelberg, Berlin, Heidelberg (2012)
13. Pellon, G., Raucent, B., Philippette, T., Mathelart, C., Alvarez, J., Kervyn, N., Cambier, F., Vangrunderbeeck, P., Motte, I., Malcourant, E., Renson, V. : Jouer pour apprendre dans l'enseignement superieur ? Play-t-il ? (Nov 2020), `https://hal.science/hal-04684462`
14. Prismatik : Mécanicartes - Wiki. https ://www.mecanicartes.com/, last accessed : 2024-12-22
15. Slimani, A., Sbert, M., Boada, I., Elouaai, F., Bouhorma, M. : Improving Serious Game Design Through a Descriptive Classification : A Comparation of Methodologies (Oct 2016)
16. Tori, A.A., Tori, R., Nunes, F.D.L.D.S. : Serious Game Design in Health Education : A Systematic Review. IEEE Transactions on Learning Technologies **15**(6), 827–846 (Dec 2022). https://doi.org/10.1109/TLT.2022.3200583
17. Triandini, E., Fauzan, R., Siahaan, D.O., Rochimah, S., Suardika, I.G., Karolita, D. : Software similarity measurements using UML diagrams : A systematic literature review **8**(1), 10. https://doi.org/10.26594/register.v8i1.2248
18. VALAT, M. : Comparaison des modèles de jeux sérieux et métriques associés (2025). https://doi.org/10.57745/GJDVV9, `https://doi.org/10.57745/GJDVV9`

Alignement Sémantique d'Apprentissages Critiques par *Embeddings* Contextuels

Wissal Essalah[1], Nicolas Delestre[2], Nicolas Malandain[2] et Damien Olivier[1]

[1] LITIS, Université Le Havre Normandie, Normandie Université, France
[2] LITIS, INSA Rouen Normandie, Normandie Université, France
wissal.essalah@etu.univ-lehavre.fr

Résumé. Nous étudions l'alignement entre Apprentissages Critiques (AC) et Attendus d'Apprentissage Disciplinaire (AAD), enjeu majeur de l'approche par compétences. Le problème est formulé comme une tâche de recherche d'information : les AC sont vus comme des requêtes, les AAD comme des documents classés par similarité sémantique. Nous comparons des modèles d'embeddings statiques (Word2Vec, FastText) et dynamiques (`SBERT`, `sentence-camembert-large`). L'analyse des courbes ROC et score AUC montrent sur nos jeux de données des résultats de moyens à bons.

Mots-clé: Approche par compétences, Évaluation pédagogique, Apprentissages Critiques, Attendus d'Apprentissage Disciplinaire, Recherche d'information, Alignement sémantique, *Embeddings* statiques, *Embeddings* dynamiques

Abstract. We study the alignment between Critical Learning Outcomes (AC) and Disciplinary Learning Objectives (AAD), a key issue in competency-based education. The task is framed as an information retrieval problem : ACs are considered as queries and AADs as documents ranked by semantic similarity. We compare static (Word2Vec, FastText) and contextual (`SBERT`, `sentence-camembert-large`) embedding models. The analysis of ROC curves and AUC scores shows results ranging from average to good on our datasets.

Keywords: Competency-based approach, Educational assessment, Critical Learning Outcomes, Disciplinary Learning Objectives, Information retrieval, Semantic alignment, Static embeddings, Contextual embeddings

1 Introduction

Depuis quelques années, les organismes de tutelle des établissements d'enseignement supérieurs français incitent à passer d'une approche programme [1] à une approche par compétences (APC) [4]. Ce changement de paradigme s'inscrit dans une dynamique de renforcement des compétences clés des étudiants.

Lors des évaluations, et en jury, les enseignants doivent alors vérifier que ces compétences et leurs prérequis sont partiellement ou totalement acquis. Ces compétences, nommées aussi savoir-agir [16], sont évaluées par des projets intégratifs. Les prérequis (savoir, savoir-faire et savoir-être), nommés « apprentissages critiques » (AC), sont enseignés et évalués dans le cadre des cours. L'évaluation de ces AC pose problème car d'une part ils sont transversaux aux différentes disciplines, aux différents cours, et d'autre part ces prérequis sont définis de manière abstraite non opérationnelle. Une façon d'évaluer ces AC est de prendre en compte leurs instanciations opérationnelles, que l'on nomme ici « Attendus d'Apprentissage Disciplinaire » (AAD), au sein des cours. Dans la formation ITI de l'INSA Rouen Normandie, le tableau 1 présente deux exemples d'AC avec quelques AAD d'un cours d'algorithmique pour la compétence « spécifier ». Cela signifie que lorsqu'un étudiant valide l'AAD « Décomposer logiquement un problème », il valide en partie l'AC « Décomposer et organiser un processus ».

Apprentissages Critiques	**Attendus d'Apprentissage Disciplinaire**
Décomposer et organiser un processus	AAD1. Formaliser les préconditions d'une opération d'un Type Abstrait de Données AAD2. Définir des Type Abstrait de Données génériques AAD3. Décomposer logiquement un problème
Utiliser un langage formel	AAD1. Formaliser les préconditions d'une opération d'un Type Abstrait de Données AAD4. Formaliser des opérations d'un Type Abstrait de Données AAD5. Formaliser des axiomes ou savoir définir la sémantique d'une opération d'un Type Abstrait de Données

TABLEAU 1 – Extrait d'alignements d'AC et AAD pour la compétence « spécifier »

Deux problèmes apparaissent alors : l'identification de ces AAD au sein d'un cours et l'alignement des AAD aux AC. Dans cet article, nous traitons du problème de l'alignement des AAD avec les AC en supposant que ces AAD et AC ont été préalablement identifiés.

2 État de l'art

L'alignement des AC et des AAD peut être abordé comme un problème de « recherche d'information » (RI), dans lequel chaque AC joue le rôle de requête et les AAD celui de documents candidats à classer selon leur pertinence [9].

Les AC et les AAD sont transformés en vecteurs numériques appelés *embeddings*, permettant de calculer leur similarité. L'efficacité de cette méthode

dépend de la capacité des modèles d'*embedding* à représenter finement la sémantique des formulations textuelles [17].

2.1 *Embedding*

L'*embedding* est une représentation vectorielle utilisée en TAL pour comparer des mots, des phrases ou des documents sur le plan sémantique [17]. Ces représentations se divisent en deux grandes familles : statiques et dynamiques.

Les *embeddings* statiques associent à chaque mot un vecteur fixe, indépendamment du contexte. Word2Vec [10], FastText [2], GloVe [11] et LexVec [15] en sont des exemples. Au niveau des phrases, la moyenne des vecteurs mots est souvent utilisée. D'autres modèles comme USE [3] produisent des vecteurs de phrases fixes mais insensibles au contexte.

Les *embeddings* dynamiques produisent des vecteurs qui varient selon le contexte d'apparition. ELMo [12], BERT [6], `SBERT` [14] et T5 [13] modélisent finement les dépendances sémantiques et syntaxiques. Pour la langue française, CamemBERT et `sentence-camembert-large` offrent des représentations contextuelles adaptées aux tâches de similarité.

2.2 Mesure de la similarité

La similarité entre représentations vectorielles est fréquemment mesurée à l'aide de la similarité cosinus [8], une métrique largement utilisée en traitement automatique des langues pour comparer des *embeddings*. Elle évalue l'angle entre deux vecteurs u et v selon :

$$\text{Similarité cosinus}(u, v) = \frac{u \cdot v}{\|u\| \, \|v\|}$$

Cette mesure renvoie une valeur comprise entre -1 et 1 : plus elle est proche de 1, plus les deux vecteurs sont considérés comme sémantiquement proches.

2.3 Critères d'évaluation

La qualité du classement produit par un système de recherche d'information est évaluée à l'aide de la courbe ROC (*Receiver Operating Characteristic*) et de l'AUC (*Area Under Curve*) [5]. Une AUC proche de 1 indique que les documents pertinents sont bien classés en tête. Une AUC $>$ 0,9 est considérée comme excellente, entre 0,8 et 0,9 elle est bonne, entre 0,7 et 0,8 acceptable, et inférieure à 0,7 considérée comme faible [7].

Le tableau 2 montre un exemple de moteur de RI retournant 10 documents, dont 4 pertinents. La figure 1 présente la courbe ROC correspondante avec une AUC de 0.83.

Rang	1	2	3	4	5	6	7	8	9	10
Pertinence	1	1	0	1	0	0	1	0	0	0
x_i	0/6	0/6	1/6	1/6	2/6	3/6	3/6	4/6	5/6	6/6
y_i	1/4	2/4	2/4	3/4	3/4	4/4	4/4	4/4	4/4	4/4

TABLEAU 2 – Résultats jouets d'un moteur de RI

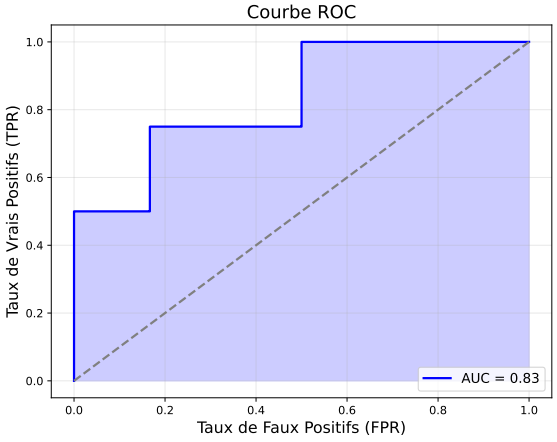


FIG. 1 – Courbe ROC correspondant aux données du tableau 2

3 Protocole expérimental

Pour valider notre approche, nous utilisons les AC définis dans la spécialité ITI de l'INSA Rouen Normandie, structurés autour de six savoir-agir (spécifier, modéliser, communiquer, concevoir, développer, gérer), chacun décomposé en quatre jalons pédagogiques. Les responsables de plusieurs cours ont identifié les AAD et réalisé un alignement avec les AC, constituant une base d'évaluation réaliste. Le tableau 3 résume les principales statistiques du corpus, incluant le nombre d'AC et d'AAD ainsi que la longueur moyenne en nombre de mots des descriptions.

AC		AAD		
Nb	**Longueur moy. (nb mots)**	**Cours**	**Nb**	**Longueur moy. (nb mots)**
79	7.54	Algorithmique	76	7.76
		Java	40	4.68
		Base de Données	44	10.89
		Électronique	20	9.10
		Python	30	6.13
		Analyse Numérique	51	7.79
		Trait. du Signal	17	5.56
		Compilation	24	10.72

TABLEAU 3 – Statistiques descriptives du corpus

3.1 Méthodologie

Notre méthodologie repose sur quatre étapes principales, illustrées dans la figure 2.

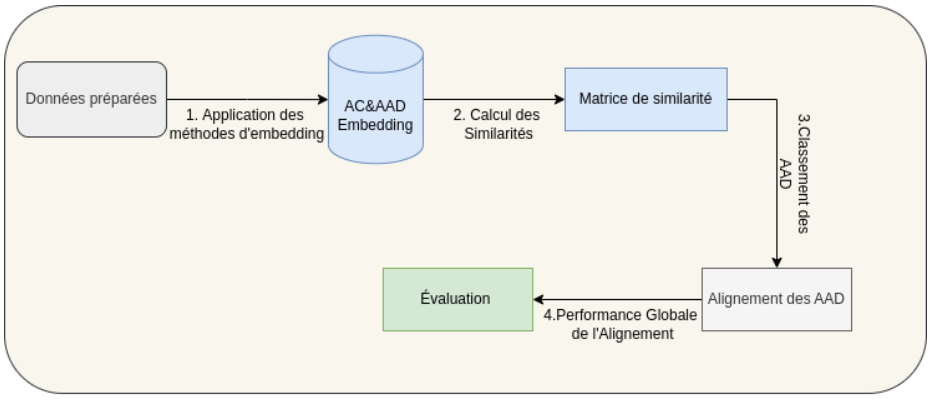


FIG. 2 – Diagramme des étapes de notre méthodologie adoptée.

Étape 1 : Application des méthodes d'Embedding. Quatre modèles issus de la section 2.1 ont été sélectionnés.

Pour les *embeddings* statiques, Word2Vec et FastText ont été appliqués après tokenisation. Word2Vec utilise l'algorithme Skip-Gram avec une dimension vectorielle de 300 et une fenêtre contextuelle de taille 5. FastText, configuré avec une dimension de 100 et une fenêtre de 5, intègre des *n-grammes* pour mieux gérer les variations morphologiques.

Les représentations de phrases ont été obtenues en moyennant les vecteurs des mots.

Pour les *embeddings* dynamiques, nous utilisons `SBERT` (préentraîné sur des données multilingues) et `sentence-camembert-large`, un modèle français optimisé pour la similarité de phrases. Ces modèles produisent des vecteurs contextuels de dimension 768.

Étape 2 : Calcul des similarités. Une matrice de similarité est construite, chaque ligne représentant un AC, chaque colonne un AAD. Les scores sont calculés à l'aide de la similarité cosinus et représentent la proximité sémantique entre les deux éléments. Un extrait de cette matrice est présenté dans le tableau 4.

Étape 3 : Classement des AAD. Pour chaque AC, les AAD sont triés selon leur score de similarité décroissant. Ce classement permet de déterminer les alignements potentiellement pertinents.

Étape 4 : Évaluation des performances. À chaque couple (AC, AAD), une vérité terrain (1 si aligné, 0 sinon) est associée. Ces annotations sont comparées au classement généré pour construire les courbes ROC et calculer les scores AUC. L'AUC moyen est ensuite utilisé pour évaluer globalement chaque modèle. Un histogramme des AUC individuels permet enfin d'analyser les performances selon les AC.

AC	**AAD1**	**AAD2**	**AAD3**	**AAD4**	**AAD5**
specifier-1.1 ("Décomposer et organiser un processus")	0.70	0.67	0.60	0.24	0.31
specifier-1.3 ("Utiliser un langage formel")	0.62	0.19	0.43	0.74	0.70

TABLEAU 4 – Extrait de matrice de similarité entre les AC "specifier-1.1" et "specifier-1.3" (voir tableau 1) et leurs AAD associés.

4 Résultats

Les performances des différentes méthodes d'*embedding* sont résumées dans le tableau 5, qui présente les scores AUC moyens obtenus pour huit cours.

On observe que les modèles dynamiques surpassent systématiquement les modèles statiques. En particulier, `SBERT`, utilisé ici dans sa version multilingue, obtient les meilleures performances dans les cours d'Algorithmique (0.79), Java (0.69) et Base de Données 0.63. Le modèle `sentence-camembert-large`, spécifiquement conçu pour le français, surpasse `SBERT` dans plusieurs cas : il atteint les meilleures performances en Électronique (0.82), en Analyse Numérique (0.62), en Traitement du Signal (0.61), en Compilation (0.72) et à égalité en Python (0.67).

Ces résultats confirment l'intérêt des *embeddings* dynamiques, et plus précisément des modèles contextualisés, pour capturer des nuances sémantiques fines dans les descriptions courtes et techniques des AC et des AAD. En comparaison,

Cours	Embeddings Statiques		Embeddings Dynamiques	
	Word2Vec	**FastText**	**sbert**	**Sentence-camembert-large**
Algorithmique	0.54	0.55	**0.79**	0.74
Java	0.49	0.56	**0.69**	0.65
Base de données	0.49	0.51	**0.63**	0.60
Électronique	0.48	0.54	0.73	**0.82**
Python	0.54	0.59	0.67	**0.67**
Analyse numérique	0.52	0.586	0.60	**0.62**
Traitement du signal	0.45	0.45	0.49	**0.61**
Compilation	0.56	0.55	0.63	**0.72**

TABLEAU 5 – Scores AUC moyens obtenus pour les embeddings statiques et dynamiques dans huit cours.

les modèles statiques comme Word2Vec et FastText produisent des résultats nettement plus faibles, avec des scores AUC généralement inférieurs à 0.60.

La figure 3 présente une comparaison des scores AUC pour chaque AC du cours d'Électronique, obtenus avec les modèles `SBERT` et `sentence-camembert-large`. Les résultats montrent une supériorité systématique de `sentence-camembert-large`, notamment pour *specifier-1.1* (0, 91 contre 0, 67), *modeliser-1.2* (0, 96 contre 0, 87) ou encore *developper-1.1* (0, 90 contre 0.75).

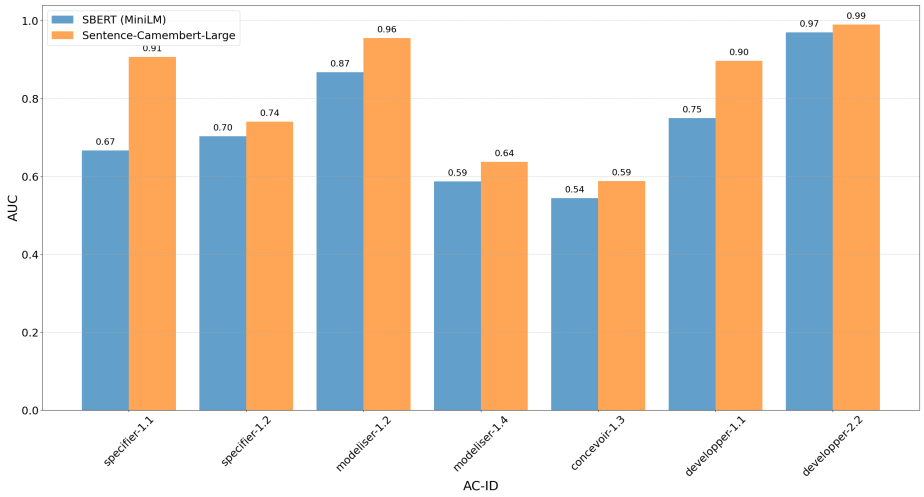


FIG. 3 – Comparaison des scores AUC pour chaque AC du cours Électronique avec `SBERT` et `sentence-camembert-large`.

Ces variations peuvent être attribuées à plusieurs facteurs. Le premier facteur concerne le nombre d'AAD associés à un AC par les responsables de cours. Par exemple, l'AC *developper-2.2* n'est associé qu'à deux AAD par l'enseignant. Le second facteur, qui est une hypothèse, concerne l'aspect explicite des AC et des AAD : le nombre de mots, l'implicite, la polysémie, etc.

5 Conclusion

Nous avons proposé une méthode d'alignement sémantique entre les AC et les AAD, dans le cadre de l'Approche Par Compétences. Le problème est formulé comme une tâche de recherche d'information, fondée sur des *embeddings* vectoriels.

L'évaluation menée sur huit cours montre la supériorité des modèles d'*embedding* dynamiques, notamment `SBERT` et `sentence-camembert-large`. Ces *embeddings* semblent être capables de capturer des relations contextuelles complexes, malgré des descriptions courtes et un déséquilibre important entre AC et AAD.

Comme perspectives, nous prévoyons de reformuler certains AC et AAD pour être plus explicite et donc améliorer la qualité des alignements. Cette hypothèse fait l'objet de premiers travaux encourageants.

Références

1. Basque, J., Rogozan, D. : Vers une méthode d'ingénierie de l'approche-programme en enseignement supérieur (2013)
2. Bojanowski, P., Grave, E., Joulin, A., Mikolov, T. : Enriching word vectors with subword information. In : Transactions of the Association for Computational Linguistics. vol. 5, pp. 135–146 (2017)
3. Cer, D., Yang, Y., Kong, S.y., Hua, N., Limtiaco, N., St. John, R., Constant, N., Guajardo-Cespedes, M., Yuan, S., Tar, C., et al. : Universal sentence encoder. In : Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing : System Demonstrations. pp. 169–174 (2018)
4. Chauvigné, C., Coulet, F. : L'approche par compétences dans l'enseignement supérieur : quelles fonctions pour l'enseignant ? (2010)
5. Chevallet, J.P. : Recherche d'information. Eyrolles (2013)
6. Devlin, J., Chang, M.W., Lee, K., Toutanova, K. : Bert : Pre-training of deep bidirectional transformers for language understanding. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics : Human Language Technologies pp. 4171–4186 (2019)
7. Şeref Kerem ÇORBACIOĞLU, AKSEL, G. : Receiver operating characteristic curve analysis in diagnostic accuracy studies : A guide to interpreting the area under the curve value. Turkish journal of emergency medicine (Online) **23**(4) (2023)
8. Lahitani, S., et al. : Cosine similarity as a measure of textual proximity in information retrieval systems. International Journal of Information Retrieval Research (2016)

9. Manning, C.D., Raghavan, P., Schütze, H. : Introduction to Information Retrieval. Cambridge University Press (2008)
10. Mikolov, T., Chen, K., Corrado, G., Dean, J. : Efficient estimation of word representations in vector space (2013)
11. Pennington, J., Socher, R., Manning, C.D. : Glove : Global vectors for word representation. Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP) pp. 1532–1543 (2014)
12. Peters, M.E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., Zettlemoyer, L. : Deep contextualized word representations. arXiv preprint arXiv :1802.05365 (2018)
13. Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, T., Matena, M., Zhou, Y., Li, W., Liu, P.J. : Exploring the limits of transfer learning with a unified text-to-text transformer. Journal of Machine Learning Research **21**(140), 1–67 (2020)
14. Reimers, N., Gurevych, I. : Sentence-bert : Sentence embeddings using siamese bert-networks. Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing pp. 3982–3992 (2019)
15. Salle, A., Villavicencio, A., Idiart, M. : Matrix factorization with additional constraints for improved word representations. In : Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics : Volume 2, Short Papers. pp. 230–236 (2016)
16. Tardif, J. : L'évaluation des compétences : Documenter le parcours de développement (2006)
17. Wang, S., Zhou, W., Jiang, C. : A survey of text representation and embedding techniques in nlp. IEEE Access **8**, 69617–69637 (2020)

Étude exploratoire de la pertinence des traces pour l'évaluation d'un projet de programmation

Christopher Marshall-Breton[1][0009−0007−2204−9613], Maude Pupin[1][0000−0003−3197−0715], Yvan Peter[1][0000−0002−1145−357X] et Mirabelle Nebut[1]

Univ. Lille, CNRS, Centrale Lille, UMR 9189 CRIStAL, F-59000 Lille, France
christopher.marshallbreton.etu@univ-lille.fr

Résumé. L'évaluation des acquis de programmation peut, en partie, être automatisée. De nombreux outils existent mais présentent cependant deux limites majeures : leur transférabilité à d'autres contextes pédagogiques, et la pertinence pour l'évaluation d'un projet. Nous avons donc un ensemble de traces remontées par ces différents outils et une difficulté à les appliquer à l'évaluation des projets. L'objet de cet article est de proposer une étude exploratoire pour étudier la faisabilité d'une évaluation entièrement automatique pour un projet de programmation. Nous nous concentrerons sur les débutants en programmation.

Mots-clé: Évaluation · Outils d'évaluation automatique · Projet de programmation · Traces d'apprentissage · Enseignement de la programmation.

Abstract. The evaluation of programming knowledge can be partly automated. Many tools exist but have two major limitations: Their transferability to other educational contexts, and the relevance for the evaluation of a project. We therefore have a set of traces returned by these different tools and a difficulty in applying them to the evaluation of projects. The purpose of this article is to propose an exploratory study to study the feasibility of a fully automatic evaluation for a programming project. We will focus on beginners.

Keywords: Evaluation · Automated Assessment Tools · Developpment Project · Learning Analytics · Programming Teaching.

1 Introduction

D'après FranceTravail [4], les besoins dans le numérique sont toujours en augmentation. Le métier de développeur, en 2025, reste le plus en tension. Dans les formations post-bac, l'informatique attire, avec 10.5% des inscriptions en cycle ingénieur dans le domaine informatique [9], et les promotions ont des effectifs conséquents. Ce nombre d'étudiants entraîne deux conséquences : D'une part,

l'évaluation devient une activité chronophage, d'autre part, de multiples parcours se développent pour les mêmes objectifs de formation. Une des réponses au premier point repose sur des outils automatiques de correction. Plusieurs outils et études se sont intéressés à ces pratiques dans un cadre d'évaluation [1,10,8]. Cependant, ces outils présentent deux limitations majeures :

Ils sont d'abord liés au processus d'apprentissage (langage, concepts évalués...). La transférabilité à un nouveau contexte pédagogique est donc difficile, et pousse souvent les équipes à préférer développer un nouvel outil propre à leurs contextes, entraînant une multiplication du nombre d'outils, avec une portée limitée [11]. Ce phénomène est exacerbé par le nombre de formations différentes engendrées par le nombre important d'étudiants comme mentionné ci-dessus.

La seconde limitation identifiée est l'inadéquation apparente de ces outils pour l'évaluation des projets. En effet, les contraintes liées à l'automatisation de l'évaluation s'opposent à la créativité recherchée dans ces projets.

L'objet de cet article est de proposer une étude exploratoire pour examiner dans quelle mesure les analyses de traces utilisées par ces outils restent pertinentes sur des projets. On étudiera également la transférabilité de ces traces à un cadre de formation différent. Nous utiliserons pour 'traces' la définition de Learning Analytics établie pour la conférence LAK 2011 et adoptée par la Society for Learning Analytics Research (SoLAR) [3], à savoir la mesure, la collecte, l'analyse et la communication de données sur les apprenants et leurs contextes, dans le but de comprendre et d'optimiser l'apprentissage et les environnements dans lesquels il se déroule.

2 Outils existants et traces

L'évaluation des étapes d'apprentissage de la programmation est complexe. Les notions à enseigner sont difficiles à identifier ou restent discutées quand elles le sont [7,13]. Aux premiers pas de l'apprentissage, on se concentre sur des exercices simples, isolés, avec un focus sur le résultat. Plus on progresse, plus les critères de validation se multiplient. Généralement, le processus d'apprentissage présente les différents concepts de la programmation par des exercices ciblés, puis permettent aux étudiants de les mettre en pratique par des projets plus larges.

De nombreux outils d'évaluation automatique existent, se concentrant majoritairement sur une succession d'exercices évalués par tests unitaires. Ces exercices sont propres à une progression pédagogique et sont généralement liés au cadre d'enseignement (utilisation d'un langage donné, notions absentes de certains exercices car pas encore abordées, *etc.*) [1,10,8]. Ce cadre restreint permet une évaluation automatique plus simple, car ces exercices sont justement très cadrés.

De ces exercices, on peut tirer un ensemble de traces de natures très différentes. L'analyse statique, sans exécution de code, permet par exemple d'avoir des informations sur le respect des conventions de nommage, sur la complexité cyclométrique, le nombre de commentaires...). L'analyse dynamique permet quant à elle d'avoir de informations sur l'exécution du code (validation de tests uni-

taires, complexité temporelle ou spatiale, erreurs de compilation/exécution...). Les traces que l'on peut remonter dépendent évidement de l'outil utilisé. C. Douce *et al.* [2] classent les outils d'évaluation automatique selon trois générations : les systèmes d'évaluation précoces (dès 1960) donnant des informations rudimentaires sur l'exécution du programme et difficiles à mettre en place, les systèmes axés sur des outils qui élargissent les possibilités de tests automatiques, augmentent le nombre de métriques mesurées et facilite le déploiement, puis les systèmes orientés web. Ces derniers, en plus de faciliter la soumission des travaux et de permettre un suivi plus personnalisé, offrent un contrôle plus précis sur le contexte d'évaluation, ce qui permet d'isoler des traces plus variées (temps passé sur un code, touches pressées...).

3 Projets de programmation

Dans une séquence pédagogique classique, les exercices sont suivis par un ou plusieurs projets, qui permettent une mise en pratique plus libre des différentes notions apprises. Les pédagogies par projet sont aujourd'hui très largement étudiées et appliquées [5]. Dans l'apprentissage de la programmation, ces projets permettent aux étudiants de mobiliser leurs acquis avec une plus grande autonomie que celle limitée par le cadre restreint des exercices isolés.

Pourtant, la liberté offerte par ces projets rend difficile leur évaluation automatique. Celle-ci se fait donc généralement manuellement par l'enseignant, par revue de code, soutenance, ou rapport.

Les projets sont également essentiels par la mise en situation et l'autonomie nécessaire à leur réalisation. Plus que la simple restitution des éléments appris, ils permettent souvent de confronter les étudiants à une situation nouvelle, et d'évaluer comment ils y font face (par exemple : utilisation d'une bibliothèque graphique sans explication préalable).

Ces projets permettent enfin de se rapprocher du contexte professionnel. Ils permettent notamment de présenter des concepts plus larges, qui n'ont que peu de sens pour des exercices isolés (gestionnaire de version, batterie de tests, intégration...)

Nous proposons donc la question de savoir si un ensemble de traces remontées dans un cadre restreint, reste pertinent sur un cadre plus large (projet) et si l'on peut étendre ce projet à une formation différente.

4 Transposition des traces à un projet : cadre d'étude

L'objectif premier de cette étude est d'identifier les traces les plus pertinentes. Pour cela, il est nécessaire de produire des traces de natures différentes.

De nombreux outils d'évaluation automatique existent, les plus connus étant Online Judge [12], ou CodingGame pour le secteur privé. Comme précisé plus haut, le nombre d'outils existant rend difficile le choix d'un en particulier, chacun ayant ses spécificités [14].

En tant que première expérience, nous travaillerons avec l'outil que notre public utilise actuellement. Il s'agit d'un plugin de l'éditeur Thonny, développé par l'Université de Lille.[6] Nous utiliserons également le langage pratiqué par les étudiants pendant ce semestre : Python.

Les traces collectées seront celles permises par l'outil, à savoir en majorité des tests fonctionnels, des informations sur la temporalité de la création des fonctions, ou de l'analyse statique sur le code rendu.

La réflexion se fera surtout sur l'analyse de ces traces. Nous voulons donc d'une part savoir quelles traces peuvent être facilement remontées, d'autre part nous intéresser à l'analyse et à la pertinence de celles-ci dans un contexte plus large.

Prenons un exemple plus concret : Une génération de docstring. Pour un exercice isolé simple, nous pouvons extraire la docstring d'une fonction unique à signature imposée demandée par l'énoncé, et vérifier sa pertinence. Dans un projet plus large et moins directif, il devient compliqué de remonter et analyser les docstrings de toutes les fonctions. La question devient donc : comment peut-on adapter le projet pour conserver la pertinence de cette trace, ou est-elle inutile car trop générale et donc difficilement exploitable ? Cette situation s'applique à de nombreuses autres traces, ayant chacune des problématiques, d'où l'objectif de cette étude. Nous avons déjà identifié plusieurs difficultés, l'étude vise à les contourner.

Notons également que nous jugeons la "pertinence" sur deux axes : La trace est-elle exploitable dans un contexte projet ? La trace permet-elle de juger d'une compétence ?

5 Construction du projet

Pour des raisons logistiques et de cohérence avec le programme de l'année, l'évaluation durera 45 min. Nous aurons 3 exercices. Un exercice de compréhension de code, nécessitant une amélioration, un exercice de recherche d'erreur nécessitant une correction, et un exercice d'écriture de code nécessitant de produire le corps d'une fonction.

Le but de cette évaluation est de se séparer d'une organisation classique d'exercices indépendants. Pour se rapprocher d'une situation professionnelle, un contexte a été imaginé : l'étudiant est en stage dans une entreprise de développement et reçoit des mails de son manager, lui demandant de réaliser certaines tâches.

Chaque exercice est présenté comme une de ces tâches. Il contient en en-tête le mail commenté, puis un morceau de code. Contrairement aux exercices de l'année, le code fourni contient volontairement des oublis de bonnes pratiques. Il contient par exemple un oubli dans les annotations de types, ou une absence de cohérence dans le nom des fonctions (franglais).

Nous ne détaillerons pas ici les différents exercices, mais nous pouvons lister quelques traces que nous pouvons en extraire (Tableau 1).

Tableau 1. Exemples de traces remontées par les 3 exercices.

N°	Descriptif de l'exercice	Objectif général	Traces
1	Ajout d'option à un code préexistant	Mise à niveau de code, Compréhension de code	Fonctionnel (Test unitaire), Mise à jour documentation (Analyse statique), Homogénéisation du code (Analyse statique)
2	Correction d'erreur dans un code préexistant	Correction de code, Compréhension de code	Fonctionnel sur borne de boucle (Test unitaire), Fonctionnel sur ré initialisation de compteur (Test unitaire), Optimisation de code (Analyse statique)
3	Écriture d'une fonction dans un environnement préexistant	Écriture de code	Navigation dans les fichiers (Logs), Temps de résolution (Logs), Utilisation des données test (Analyse statique), Fonctionnel (Test unitaire)

Cette base nous permettra dans un second temps de mettre en place une épreuve projet beaucoup plus large, afin d'étudier les traces isolées dans les exercices précédents dans un contexte de projet.

6 Conclusion

Les outils d'évaluation automatique sont nombreux et sont souvent limités à un contexte d'enseignement très particulier. C'est pourtant ce cadre restreint qui permet de tirer relativement simplement diverses traces d'apprentissage à partir des exercices d'étudiants.

Nous allons dans un premier temps récupérer un ensemble varié de traces à partir d'exercices simulant une situation professionnelle. Ces traces seront ensuite analysées pour évaluer leur pertinence dans un contexte plus large, à savoir un projet de programmation.

Remerciements. This work is funded by the Interreg France-Wallonie-Vlaanderen *Open Badges for IT* project.

References

1. Combéfis, S.: Automated Code Assessment for Education: Review, Classification and Perspectives on Techniques and Tools. Software **1**(1), 3–30 (Feb 2022). `https://doi.org/10.3390/software1010002`, `https://www.mdpi.com/2674-113X/1/1/2`
2. Douce, C., Livingstone, D., Orwell, J.: Automatic test-based assessment of programming: A review. J Educ Resour Comput **5** (Sep 2005). `https://doi.org/10.1145/1163405.1163409`

3. Ferguson, R.: Learning analytics: drivers, developments and challenges. International Journal of Technology Enhanced Learning **4**(5-6), 304–317 (2012)
4. France-Travail: Les métiers du numérique : Tour d'horizon des enjeux et des opportunités, https://www.francetravail.org/accueil/actualites/infographies/2025/les-metiers-du-numerique-tour-d'horizon-des-enjeux-et-des-opportunites.html?type=article
5. Kokotsaki, D., Menzies, V., Wiggins, A.: Project-based learning: A review of the literature. Improving schools **19**(3), 267–277 (2016)
6. Marvie-Nebut, M., Peter, Y.: Apprentissage de la programmation python-une première analyse exploratoire de l'usage des tests. In: Atelier «Apprendre la Pensée Informatique de la Maternelle à l'Université», dans le cadre de la conférence Environnements Informatiques pour l'Apprentissage Humain (EIAH). pp. 1–8 (2023)
7. Medeiros, R.P., Ramalho, G.L., Falcão, T.P.: A Systematic Literature Review on Teaching and Learning Introductory Programming in Higher Education. IEEE Transactions on Education **62**(2), 77–90 (May 2019). https://doi.org/10.1109/TE.2018.2864133, https://ieeexplore.ieee.org/document/8447543/?arnumber=8447543, conference Name: IEEE Transactions on Education
8. Messer, M., Brown, N.C.C., Kölling, M., Shi, M.: Automated Grading and Feedback Tools for Programming Education: A Systematic Review. ACM Transactions on Computing Education **24**(1), 1–43 (Mar 2024). https://doi.org/10.1145/3636515, https://dl.acm.org/doi/10.1145/3636515, publisher: Association for Computing Machinery (ACM)
9. Ministere-ESR: Les effectifs inscrits en cycle ingénieur en 2023-2024 (Jun 2024), https://www.enseignementsup-recherche.gouv.fr/fr/les-effectifs-inscrits-en-cycle-ingenieur-en-2023-2024-96520
10. Paiva, J.C., Leal, J.P., Figueira, A.: Automated Assessment in Computer Science Education: A State-of-the-Art Review. ACM Transactions on Computing Education **22**(3), 1–40 (Sep 2022). https://doi.org/10.1145/3513140, https://dl.acm.org/doi/10.1145/3513140
11. Pettit, R., Prather, J.: Automated assessment tools: too many cooks, not enough collaboration. J. Comput. Sci. Coll. **32**(4), 113–121 (Apr 2017)
12. Revilla, M.A., Manzoor, S., Liu, R.: Competitive learning in informatics: The uva online judge experience. Olympiads in Informatics **2**(10), 131–148 (2008)
13. Robins, A., Rountree, J., Rountree, N.: Learning and Teaching Programming: A Review and Discussion. Computer Science Education **13**(2), 137–172 (Jun 2003). https://doi.org/10.1076/csed.13.2.137.14200, https://www.tandfonline.com/doi/full/10.1076/csed.13.2.137.14200
14. Wasik, S., Antczak, M., Badura, J., Laskowski, A., Sternal, T.: A survey on online judge systems and their applications. ACM Computing Surveys (CSUR) **51**(1), 1–34 (2018)

Quels niveaux d'assistance l'IA peut-elle offrir aux enseignants pour faciliter le processus de création d'activités multimodales en réalité mixte ?

Mohammed Oussama Seddini[1], Mohamed Ezzaouia[2], Iza Marfisi[1]

[1] LIUM, Université du Mans, Le Mans, France
mohammed_oussama.seddini.etu@univ-lemans.fr, iza.marfisi@univ-lemans.fr
[2] IRISA, Université de Rennes
mohamed.ez-zaouia@univ-rennes.fr

Résumé. La Réalité Mixte (RM) combine éléments réels et virtuels, offrant de nouvelles opportunités éducatives. Des outils auteurs permettent désormais aux enseignants de créer des activités de RM sans compétences en programmation, mais la conception reste complexe, notamment pour la création de contenu multimédia. Cet article propose quatre niveaux d'assistance d'IA pour aider les enseignants. Nous développons une preuve de concept en intégrant des modules d'IA dans un outil auteur existant et menons une étude préliminaire avec un enseignant pour évaluer leur pertinence et utilité.

Mots-clés : Outil Auteur, Intelligence Artificielle Générative, Grands Modèles de Langage, Réalité Mixte, Education

Abstract. Mixed Reality (MR) combines real and virtual elements, offering new educational opportunities. Recently, authoring tools emerged to help teachers create MR activities without programming skills. However, designing these activities, especially creating multimodal content, remains challenging. To address this, we propose four levels of AI assistance for MR in education. We developed a proof of concept by integrating AI modules into an existing authoring tool and conducted a preliminary study with a teacher to evaluate the relevance and usefulness of our approach.

Keywords: Authoring tool, Generative Artificial Intelligence, Large Language Modeling, Mixed Reality, Education

1 Introduction

La Réalité Mixte (RM) associe objets réels et virtuels, facilitant l'apprentissage de concepts complexes [1] et offrant un apprentissage multisensoriel bénéfique dans diverses disciplines. Bien que la RM stimule l'engagement et la mémorisation [2], sa création reste complexe pour les enseignants en raison des compétences de programmation nécessaires (Unity, Vuforia, ARCore) [3, 4, 5]. Des outils comme MIXAP offrent des solutions sans programmation [6, 7], mais des défis persistent, notamment pour trouver des idées pédagogiques adaptées [8]. L'IA générative pourrait surmonter ces limites en facilitant la création de contenu sans expertise technique [9].

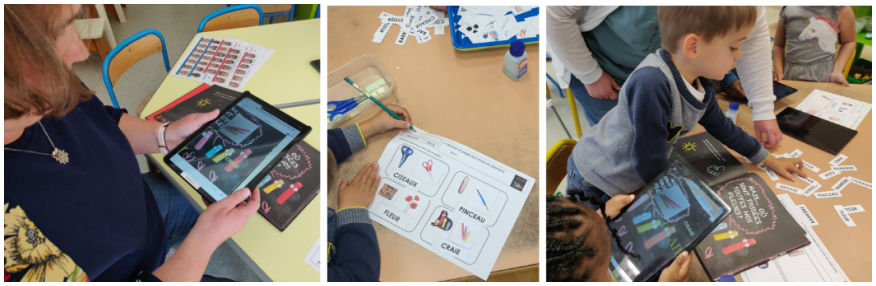


Fig. 1. Une professeure des écoles, qui utilise un outil auteur pour créer une activité de lecture collaborative intégrant de la Réalité Mixte [7].

Dans cet article, nous explorons comment l'IA peut faciliter la création d'activités pédagogiques en RM, en répondant à la question suivante :

- **RQ1 : Quels niveaux d'assistance l'IA peut-elle apporter aux enseignants pour leur faciliter le processus de création d'activités pédagogiques intégrant de la RM ?**

2 Etat de l'art

2.1 Outils auteurs d'activités pédagogiques intégrant de la Réalité Mixte

En 2022, un état de l'art a recensé 21 outils auteurs pour créer des activités pédagogiques en Réalité Mixte ou Augmentée, mais aucun ne répond pleinement aux besoins des enseignants. Beaucoup nécessitent une connexion Internet, des tablettes performantes, et une formation. MIXAP, un outil auteur open source, facilite la création d'activités en RM. Il propose des modèles prédéfinis, comme l'augmentation d'image, avec une interface intuitive. Nous avons choisi d'intégrer l'assistance IA dans MIXAP, particulièrement pour l'étape 3, où les enseignants ajoutent des augmentations sur les ressources pédagogiques [10][11].

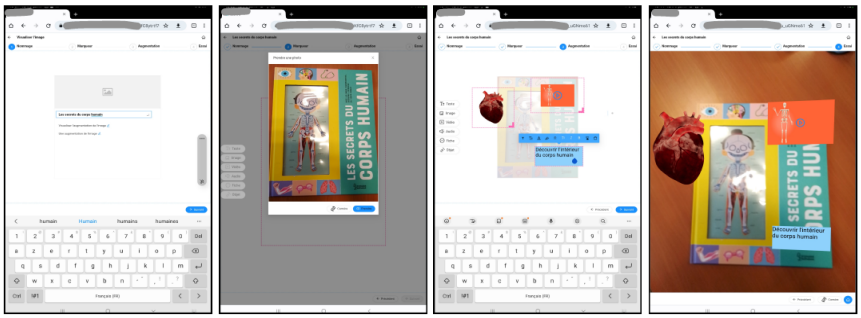


Fig. 2. Processus de conception proposé par MIXAP

2.2 Co-création de contenu avec des outils d'Intelligence Artificielle Générative

Les outils d'IA générative, comme GPT-4, génèrent textes, images, vidéos et code à partir de prompts en langage naturel, mais présentent des limites de précision et de cohérence [12]. L'ingénierie de prompts est essentielle pour maximiser leur potentiel, avec des formations spécifiques [13]. L'adoption croissante des modèles de traitement du langage naturel soulève des questions sur l'évolution du rôle des enseignants, passant de créateurs à concepteurs de prompts [14]. De plus, les chatbots éducatifs enrichissent les pratiques pédagogiques en servant d'assistants et de supports éducatifs, redéfinissant l'interaction humain-IA générative [15].

2.3 L'IA comme module d'assistance in-situ

Les modules d'assistance intégrant de l'IA sont de plus en plus répandus, facilitant la rédaction (Gmail), la création de contenu (LinkedIn, Canva) et l'éducation. Des outils comme PATHWiSE et AIIA offrent des leçons personnalisées et réduisent la charge cognitive [16, 17], tandis que Mimisbrunnur [21] et CodeAid [18] proposent des aides pour l'écriture interactive et la programmation. Les chatbots éducatifs utilisant des modèles LLM et RAG fournissent des réponses contextuelles [19], et des systèmes comme ChatGPT automatisent les évaluations [20]. Cependant, l'usage de l'IA pour créer des activités pédagogiques en réalité mixte (RM) reste peu exploré, malgré des défis spécifiques liés à l'alignement physique-virtuel.

3 Assistance d'IA pour la création d'activités multimodales pédagogiques en MR

3.1 Proposition de quatre niveaux d'assistance

À partir de l'état de l'art et du jeu de données (JDD) regroupant 18 activités pédagogiques en RM conçues par 18 enseignants aux profils variés (technophiles ou non, différentes matières et niveaux d'enseignement) [22], nous proposons quatre niveaux d'assistance où l'IA générative pourrait faciliter la création d'activités pédagogiques en RM. Certains enseignants du JDD seront désignés par leurs initiales.

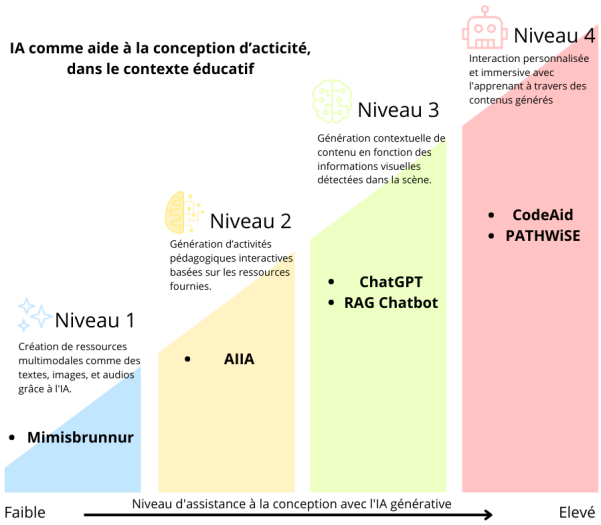


Fig. 3. Les 4 niveaux d'assistance à la conception d'activités pédagogiques avec l'IA générative.

Niveau 1: Aide à la création de ressources multimodales. Ce niveau consiste à créer des ressources multimodales (images, audio, vidéos, modèles 3D), utiles pour la mémorisation et l'engagement des apprenants [23, 24]. Les enseignants passent beaucoup de temps à chercher des ressources gratuites et sans licence. L'IA, comme dans Mimisbrunnur [21], pourrait faciliter ce processus avec des fonctions de génération audio (text-to-speech) et d'images via prompts. Par exemple, l'enseignant TN pourrait créer des enregistrements pour les élèves en difficulté de lecture. Cependant, intégrer ces outils IA dans les auteurs de RM reste un défi.

Niveau 2 : Aide à la génération d'activités de MR. vise à générer des activités de Réalité Mixte (RM) avec l'IA, en s'inspirant des travaux sur AIIA [17]. L'IA peut créer des activités complètes en utilisant des ressources pédagogiques (livre, affiche, fiche) et des objectifs d'apprentissage. Par exemple, pour l'enseignante DelD, l'IA pourrait

détecter les objets sur une page et générer automatiquement des étiquettes, images et fichiers audio, réduisant ainsi le temps de création.

Niveau 3 : Proposition de contenu en fonction des éléments détectés dans la scène. Ce niveau propose du contenu basé sur les éléments détectés dans la scène, permettant de créer des activités pédagogiques dynamiques et contextuelles [19, 20]. Contrairement aux contenus statiques des niveaux précédents, l'IA utilise la caméra pour identifier les ressources pédagogiques et suggérer des interactions adaptées. Par exemple, pour l'enseignante CP en Histoire-Géographie, l'IA pourrait détecter des cartes et images de la Révolution française et proposer des activités telles que la création d'une frise chronologique ou le placement d'événements sur la carte.

Niveau 4 : Interaction personnalisée et immersive avec l'apprenant à travers des contenus générés. Ce niveau propose des interactions personnalisées et immersives avec l'apprenant via des contenus générés [16, 18, 24]. L'IA permet de créer des expériences d'apprentissage adaptées aux niveaux et aux intérêts des apprenants, ce qui améliore leur réussite. Par exemple, PC, enseignant d'anglais, souhaite utiliser la RM pour présenter des anecdotes sur les membres de la famille royale d'Angleterre via des portraits interactifs. Grâce aux signaux multimodaux, l'IA pourrait permettre aux élèves d'interagir avec chaque portrait en posant des questions, recevant des réponses sous forme de texte ou d'audio, favorisant ainsi un apprentissage plus personnalisé.

3.2 Implémentation d'un démonstrateur

Pour tester les niveaux d'aide proposés, nous avons développé trois modules IA : IA Prompteur (niveaux 1 et 2), IA Chat (niveau 2) et IA Tuteur (niveaux 3 et 4). Intégrés dans MIXAP (open source), ces modules utilisent GPT-4 [26], DALL-E 3 [27] et OpenAI Whisper [28].

IA Prompteur. Ce module utilise les LLM d'OpenAI pour générer des contenus multimodaux (sons, images, textes) via des prompts en français dans MIXAP (Figure 2). Par exemple, après une photo d'un lion, l'enseignant génère automatiquement une image, audio ou texte via IA.

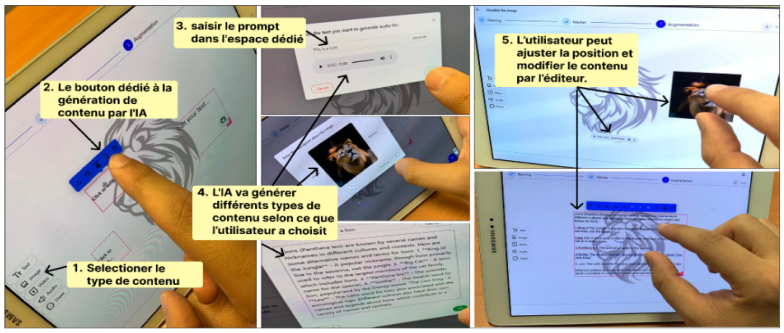


Fig. 4. Exemple d'utilisation du module AI Prompteur pour créer des ressources multimédias

IA Chat. Une fonction de chat propose des suggestions d'activités en temps réel basées sur les ressources visuelles (Figure 5). Par exemple, une photo de planisphère génère une description et des faits historiques, ajustables par l'enseignant. Ce module facilite l'idéation en proposant des contenus pédagogiques pertinents.

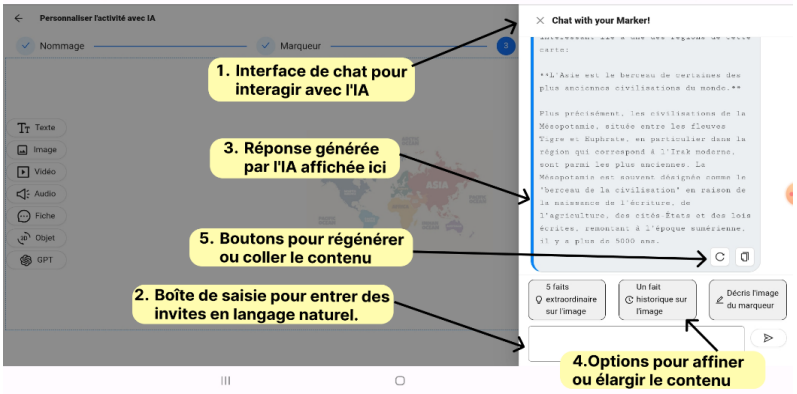


Fig. 5. Exemple d'utilisation du module AI Chat pour créer des activités

IA Tuteur. Ce module offre des expériences d'apprentissage personnalisées selon les interactions de l'élève avec la scène (Figure 6). Par exemple, en pointant un continent, l'élève reçoit des réponses contextuelles, adaptant les activités aux besoins éducatifs, comme l'apprentissage des langues ou des faits historiques.

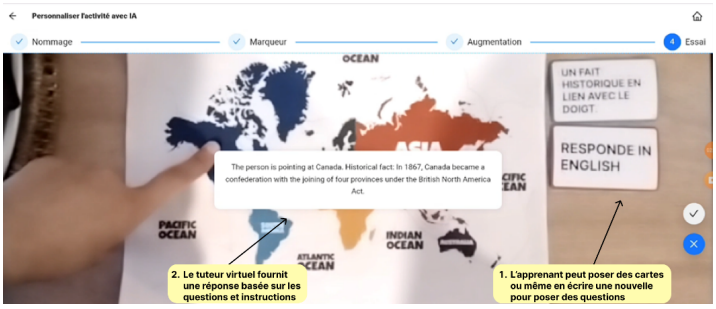


Fig. 6. Exemple d'utilisation du IA Tuteur pour créer des expériences d'apprentissage unique

4 Étude préliminaire

4.1 Méthode

Le prototype intégrant les modules d'IA confirme que les quatre niveaux d'aide sont adaptés aux activités pédagogiques en RM et répondent aux besoins identifiés dans le jeu de données [25]. Une étude préliminaire a été menée avec un enseignant de SVT expérimenté en outils numériques (Figure 6). Après une démonstration de 10 minutes de MIXAP, l'enseignant a créé quatre activités en 75 minutes, suivant des instructions prédéfinies. Les activités comprenaient l'annotation d'une imprimante 3D, l'association cartes-animaux, la géolocalisation et la validation d'étiquettes (Figures 7 et 8). La session a été enregistrée pour évaluer l'efficacité du système.



Fig. 7. Étude préliminaire avec un enseignant



Fig. 8. Matériel mis à disposition pendant l'étude préliminaire

4.2 Résultats

Le module IA Prompteur (Niveau 1) a permis de créer des contenus multimodaux (images, textes, audios) de manière rapide, mais nécessite parfois des ajustements pour plus de précision. L'enseignant a apprécié cette efficacité, tout en suggérant des améliorations pour rendre les prompts plus intuitifs. Le module IA Chat (Niveau 2) a aidé à générer des idées d'activités pédagogiques en lien avec les ressources, mais la validation humaine reste indispensable. Enfin, le module IA Tuteur (Niveaux 3 et 4) a démontré sa capacité à intégrer les éléments visuels dans l'apprentissage, bien que la reconnaissance des écritures manuscrites et des messages en français reste perfectible.

5 Conclusion et Perspectives

Cet article propose quatre niveaux d'aide d'IA générative pour assister les enseignants dans la création d'activités en réalité mixte (RM), implémentés dans l'outil MIXAP. Une étude préliminaire a confirmé leur potentiel pour enrichir les pratiques pédagogiques. Des recherches futures évalueront leur impact, amélioreront les modules et analyseront les enjeux techniques et éthiques, incluant la protection des données.

Remerciement

Nous remercions chaleureusement le consortium Ikigai porté par l'association Games for Citizens, la société Gamaizer ainsi que le projet FORTEIM (projet lauréat AMI CMA France 2030), les deux laboratoires LIUM et IRISA, pour leur soutien et leur collaboration. Leurs contributions ont apporté une valeur ajoutée significative à la réalisation de cette recherche.

Références

1. Akçayır, M., & Akçayır, G. (2017). Advantages and challenges associated with augmented reality for education: A systematic review of the literature. *Educational research review*, *20*, 1-11.
2. Yoon, S., Anderson, E., Lin, J., & Elinich, K. (2017). How augmented reality enables conceptual understanding of challenging science content. Journal of Educational Technology & Society, 20(1), 156-168.
3. Unity Technologies. [n. d.] Unity real-time development platform. (). https://unity.com/.
4. PTC Inc. [n. d.] Vuforia: augmented reality platform. (). https://www.ptc.com/en/products/vuforia.
5. Google. [n. d.] Arcore: augmented reality sdk for android. (). https://developers.google.com/ar.
6. Cypher, A. (2010). End user programming on the Web. In *No Code Required* (pp. 3-22). Morgan Kaufmann.
7. Ez-Zaouia, M., Marfisi-Schottman, I., & Mercier, C. (2023, June). MIXAP : Un outil auteur d'activités éducatives en réalité augmentée. In *Environnements informatiques pour l'apprentissage humain (EIAH)*.
8. Ez-Zaouia, M., Marfisi-Schottman, I., & Mercier, C. (2023, April). Authoring Tools: The Road To Democratizing Augmented Reality For Education. In *Proceedings of the 15th International Conference on Computer Supported Education* (Vol. 1, pp. 115-127).
9. Chiu, T. K. (2023). The impact of Generative AI (GenAI) on practices, policies and research direction in education: A case of ChatGPT and Midjourney. *Interactive Learning Environments*, 1-17.
10. Ez-Zaouia, M., Marfisi-Schottman, I., Oueslati, M., Mercier, C., Karoui, A., & George, S. (2022, November). A design space of educational authoring tools for augmented reality. In *International Conference on Games and Learning Alliance* (pp. 258-268). Cham: Springer International Publishing.
11. https://forge.apps.education.fr/lium/mixap
12. Ge, T., Hu, J., Dong, L., Mao, S., Xia, Y., Wang, X., ... & Wei, F. (2022). Extensible prompts for language models. *arXiv preprint arXiv:2212.00616*.
13. Bang, Y., Cahyawijaya, S., Lee, N., Dai, W., Su, D., Wilie, B., ... & Fung, P. (2023). A multitask, multilingual, multimodal evaluation of chatgpt on reasoning, hallucination, and interactivity. *arXiv preprint arXiv:2302.04023*.
14. Clegg, S., & Sarker, S. (2024). Artificial intelligence and management education: A conceptualization of human-machine interaction. *The International Journal of Management Education*, *22*(3), 101007.
15. Bozkurt, A., & Sharma, R. C. (2024). Are we facing an algorithmic renaissance or apocalypse? Generative AI, ChatBots, and emerging human-machine interaction in the educational landscape. *Asian Journal of Distance Education*, *19*(1).
16. Rahman, M. A., Felix, I. A., Shahid, U., & Michaelis, J. E. (2024, March). PATHWiSE: An AI-Assisted Teacher Authoring Tool for Creating Custom Robot-Assisted Learning Activities. In *Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction* (pp. 88-90).
17. Sajja, R., Sermet, Y., Cikmaz, M., Cwiertny, D., & Demir, I. (2024). Artificial intelligence-enabled intelligent assistant for personalized and adaptive learning in higher education. *Information*, *15*(10), 596.
18. Kazemitabaar, M., Ye, R., Wang, X., Henley, A. Z., Denny, P., Craig, M., & Grossman, T. (2024, May). Codeaid: Evaluating a class-room deployment of an llm-based programming

assistant that balanc-es student and educator needs. In Proceedings of the CHI Conference on Human Factors in Computing Systems (pp. 1-20).
19. Modran, H., Bogdan, I. C., Ursuțiu, D., Samoila, C., & Modran, P. L. (2024). LLM intelligent agent tutoring in higher education courses using a RAG approach. Preprints 2024, 2024070519.
20. Askarbekuly, N., & Aničić, N. (2024). LLM examiner: automating assessment in informal self-directed e-learning using ChatGPT. Knowledge and Information Systems, 1-18.
21. Stefnisson, I., & Thue, D. (2018, September). Mimisbrunnur: AI-assisted authoring for interactive storytelling. In Proceedings of the AAAI Conference on artificial Intelligence and Interactive Digital en-tertainment (Vol. 14, No. 1, pp. 236-242).
22. Marfisi, iza (2022), "Teachers Paper-prototyping Augmented Reality Activities for Classroom Use", Mendeley Data, V1, doi: 10.17632/jwgghwd3z8.1
23. Van Gog, T., Verveer, I., & Verveer, L. (2014). Learning from video modeling examples: Effects of seeing the human model's face. *Computers & Education*, *72*, 323-327.
24. Zhang, L., Basham, J. D., & Yang, S. (2020). Understanding the implementation of personalized learning: A research synthesis. *Educational research review*, *31*, 100339.
25. https://lium-cloud.univ-lemans.fr/index.php/s/JknAQAsrPj2p8MW/download
26. OpenAI. 2023. Gpt-4 api. Available at https://platform.openai.com/docs/models/gpt-4. (2023).
27. OpenAI. 2023. Dall·e 3. Available at https://platform.openai.com/docs/guides/images. (2023).
28. OpenAI. 2023. Whisper: speech recognition api. Available at https://platform.openai.com/docs/guides/speech-to-text. (2023).

Repenser l'intégration : le jeu comme métaphore expérientielle pour l'apprentissage

Hélène Parmentier[1], Eric Sanchez[2]

[1] TECFA-FPSE, Université de Genève
helene.parmentier@unige.ch
[2] TECFA-FPSE, Université de Genève
eric.sanchez@unige.ch

Résumé. Cet article propose un cadre théorique pour la conception de jeux éducatifs intégrés, fondé sur l'apprentissage expérientiel et la métaphore conceptuelle. Il introduit la notion de *métaphore expérientielle*, permettant une intégration cohérente entre objectifs pédagogiques et mécaniques ludiques. Trois principes de design convergent sont proposés. Ce cadre vise à guider les concepteurs vers des expériences d'apprentissage ludiques pertinentes, en dépassant les limites des approches superficielles de type *chocolate-covered broccoli*. Des perspectives d'application et de validation sont proposées.

Mots-clés : Learning game design, intégration, expérience, jeu, métaphore

Abstract. This article proposes a theoretical framework for the design of integrated educational games, grounded in experiential learning and conceptual metaphor theory. It introduces the notion of *experiential metaphor*, enabling coherent integration between educational goals and game mechanics. Three principles of convergent design are proposed. This framework aims to guide designers toward meaningful learning experiences through play, overcoming the limitations of superficial approaches such as the *chocolate-covered broccoli* effect. Perspectives for application and validation are also discussed.

Keywords: Learning game design, integration, experience, play, metaphor

1 Introduction

Le développement des jeux éducatifs suscite un intérêt croissant, porté par leur potentiel en matière d'engagement et de motivation. Toutefois, une difficulté persiste dans l'intégration cohérente des savoirs en jeu, souvent illustrée par la métaphore du *chocolate-covered broccoli* [1] : un contenu éducatif parsemé d'éléments ludiques superficiels [2].

Face à ce constat, deux questions se posent : comment concevoir un environnement d'apprentissage articulant étroitement contenus éducatifs et jeu ? Quels fondements

théoriques mobiliser pour concevoir des expériences d'apprentissage par le jeu de qualité ?

Nous proposons un cadre théorique s'appuyant sur l'apprentissage expérientiel [3] et la métaphore conceptuelle [4] pour introduire la notion de métaphore expérientielle, un outil conceptuel au service d'un design convergent, unifiant objectifs ludiques et éducatifs.

2 Les défis d'intégration des objectifs éducatifs en jeu

La réussite d'un jeu éducatif repose sur sa capacité à éviter une dissonance entre jeu et apprentissage. La cohérence entre les activités ludiques et les objectifs pédagogiques en accroît l'intérêt et l'efficacité [5]. Cette cohérence doit s'appliquer aux mécaniques mêmes du jeu, et non se limiter à son univers narratif [6]. Les mécaniques sont ici comprises comme les règles opérationnalisées du jeu, ou les méthodes employées par les joueurs pour interagir avec l'univers du jeu [7].

Habgood va plus loin en proposant le concept d'intégration intrinsèque [2] où l'apprentissage résulte directement de l'interaction avec les mécaniques ludiques. Cette approche vise à éviter la superficialité des jeux où les contenus éducatifs sont simplement "thématisés".

L'enjeu central consiste en l'intégration des objectifs pédagogiques dans les mécaniques elles-mêmes, condition nécessaire pour dépasser l'effet *chocolate-covered broccoli*. Cependant, si ces travaux soulignent l'importance de l'intégration, ils laissent ouverte la question des principes théoriques permettant de guider la conception de tels jeux.

3 Le jeu, situation adidactique d'apprentissage expérientiel

Le concept de situation adidactique [8] offre une première piste théorique sur la question. Dans une situation adidactique, l'apprenant interagit avec un milieu didactique qui génère des rétroactions adaptées. L'apprentissage découle ainsi d'un processus d'expérimentation. Ce concept s'inscrit dans la tradition de l'apprentissage expérientiel [9], où les connaissances se construisent à la manière d'une enquête active, par *interaction* avec l'environnement et *continuité* de l'expérience (p.45).

Dans le jeu vidéo, la rhétorique procédurale [10] illustre cette perspective en décrivant comme les règles transmettent du sens. Dans un jeu "persuasif" les mécaniques portent un message que le joueur interprète en manipulant et en testant le système. Ainsi, tout comme dans la situation adidactique, le joueur, confronté à un système de règles, formule des hypothèses, les teste, reçoit des rétroactions, et ajuste ses stratégies sans intervention didactique explicite. La rhétorique procédurale peut alors être comprise comme un mécanisme opératoire de la situation adidactique appliquée au game design : les mécaniques constituent alors un milieu didactique.

Dans une perspective expérientielle, l'adidacticité est fondamentale : l'interaction avec le milieu fournit les éléments nécessaires à l'analyse et à la résolution de la situation. Le jeu conçu pour l'apprentissage permet au joueur d'apprendre en interprétant activement un système de contraintes signifiantes.

4 La métaphore comme enquête

Selon Lakoff et Johnson, la compréhension de concepts abstraits s'appuie sur des métaphores conceptuelles [4], ancrées dans l'expérience concrète, sensible. La métaphore "permet de comprendre quelque chose (et d'en faire l'expérience) en termes de quelque chose d'autre" (p.15). La métaphore ne se réduit pas à une figure de style, elle structure notre compréhension du monde et nous permet de le redécrire [11].

Dans une perspective d'apprentissage expérientiel, comprendre une métaphore est un processus d'enquête au sens de Dewey [9] : le destinataire interprète la situation en mobilisant son expérience passée, le contexte et les indices disponibles pour formuler des hypothèses jusqu'à parvenir à la signification la plus pertinente.

Appliqué au game design, concevoir une expérience métaphorique, c'est structurer l'interaction ludique autour d'une analogie implicite que le joueur doit décoder à travers son expérience. La métaphore peut structurer l'expérience ludique pour soutenir la construction du savoir : elle devient un "sous-texte" que le joueur doit interpréter activement.

5 La métaphore expérientielle pour l'apprentissage

Un jeu conçu comme une métaphore expérientielle offre au joueur des indices pour formuler, tester et ajuster ses hypothèses à travers l'interaction ludique, lui permettant de "comprendre et de faire l'expérience" de ce qu'on souhaite qu'il comprenne "dans les termes d'autre chose".

Un exemple de métaphore expérientielle peut être trouvé dans Mission Télomère [12], un jeu en réalité mixte créé pour le Centre d'Information et de Prévention du Tabagisme de Fribourg. Dans une situation de jeu, les élèves se trouvent en présence d'un bouton rouge sur lequel est écrit "danger", mais la curiosité va amener l'un des joueurs à le presser. Une fois ce bouton activé, une alarme se déclenche à une fréquence croissante, ne pouvant être interrompue que par une nouvelle activation du bouton. Le jeu ne décrit pas le phénomène d'addiction, il permet de « le comprendre et d'en faire l'expérience" à travers la manière dont il est conçu.

Pour une telle métaphore, la situation d'apprentissage est adidactique (le but du jeu est différent de l'objectif d'apprentissage, qui n'est pas déclaré), expérientielle (les objectifs d'apprentissages sont traduits en règles du jeu ; l'apprenant doit faire l'expérience du jeu pour apprendre), et métaphorique : le choix de l'objectif du jeu et de ses mécaniques forme un tout cohérent qui converge avec les objectifs d'apprentissage, les mobilisant et les reflétant à la fois.

6 Vers un design convergent

Les travaux précédents portant sur les jeux pour l'apprentissage ainsi que les fondements théoriques mobilisés jusqu'ici nous amènent à proposer trois principes de conception pour un design convergent. Dans les modèles d'ingénierie, ces principes sont pertinents pour les équipes de conception dans la phase de design, une fois les objectifs d'apprentissage définis et avant la phase de développement.

- Principe d'interaction : les objectifs éducatifs doivent être intégrés prioritairement dans les règles du jeu, et non simplement dans son habillage narratif, en ligne avec les travaux précédents traitant de l'intégration, pour permettre aux joueurs d'apprendre de leur expérience du jeu.
- Principe d'adidacticité : les objectifs éducatifs ne sont pas explicités au joueur : la conception du jeu devient élaboration d'un milieu didactique qui conditionne le succès du joueur à la découverte ou à la mobilisation des savoirs visés.
- Principe de cohérence : l'opérationnalisation des deux premiers principes passe par une métaphorisation des savoirs en jeu. Les différentes dimensions du jeu (mécaniques, graphismes, narration, etc.) doivent respecter la métaphore sélectionnée pour sa pertinence en regard des objectifs d'apprentissage, du contexte et des joueurs. Il s'agit de garantir une cohérence entre la métaphore créée pour le jeu (source) et les objectifs d'apprentissage (cible).

7 Perspectives et limites

Les équipes de conception de jeux pour l'apprentissage sont multidisciplinaires, souvent pauvres en ressources et en outils. Un dispositif permettant de guider la métaphorisation des savoirs en jeu pour concevoir une expérience cohérente pourrait fournir à ces équipes le guidage méthodologique nécessaire pour créer une expérience d'apprentissage par le jeu pertinente.

La validation empirique de ces principes de conception nécessite à ce stade une analyse approfondie de jeux existants, ainsi que l'étude de sa mise en œuvre en contexte réel de conception et d'usage de jeux éducatifs. En particulier, nous souhaitons évaluer comment un guide de conception fondé sur ces principes peut aider les enseignants et les concepteurs à créer des jeux vidéo pédagogiques "intrinsèquement intégrés" [2] que nous appelons des métaphores expérientielles.

Nous avons proposé une conception du jeu éducatif comme métaphore expérientielle pour aboutir à des principes de design convergent, en nous appuyant sur l'apprentissage expérientiel [3] et la métaphore conceptuelle [4]. Considérer le jeu comme une métaphore expérientielle permet d'asseoir l'apprentissage par le jeu dans un cadre théorique pertinent duquel on peut faire émerger des principes de design permettant une intégration convergente des savoirs en jeu, en dépassant leur apparente opposition pour permettre une expérience d'apprentissage par le jeu cohérente et pertinente.

References

1. A. Bruckman, “Can Educational Be Fun?”, in *1999 Game Developers Conf. Proc. (GDC 1999)*, 1999, pp. 75-79.
2. J. Habgood, “The Effective Integration of Digital Games and Learning Content”, Ph.D dissertation, University of Nottingham, UK, 2007.
3. J. Dewey, *Experience And Education*. Free Press, 1938.
4. G. Lakoff and M. Johnson, *Les métaphores dans la vie quotidienne*. Éd. de Minuit, 1985.
5. T. W. Malone, “Toward a theory of intrinsically motivating instruction”, *Cognitive Science*, vol. 5, no. 4, pp. 333-369, Oct. 1981, doi: 10.1016/S0364-0213(81)80017-1.
6. C. Fabricatore, “Learning and videogames: An unexploited synergy”, presented at the Annual Convention of the Association for Educational Communications and Technology (AECT 2000), Long Beach, CA, Feb. 2000.
7. M. Sicart, “Defining Game Mechanics”, *Game Studies*, vol. 11, no. 3, Dec. 2011. [Online]. Available: https://gamestudies.org/1103/articles/sicart
8. G. Brousseau, “Des dispositifs Piagétiens… aux situations didactiques”, *Éducation et didactique*, no. 6-2, pp. 103-129, Oct. 2012, doi: 10.4000/educationdidactique.1475.
9. J. Dewey, *Logic: The Theory Of Inquiry*. New York, USA: Henry Holt, 1938.
10. I. Bogost, *Persuasive Games: The Expressive Power of Videogames.* The MIT Press, 2007. doi: 10.7551/mitpress/5334.001.0001.
11. P. Ricœur, *La métaphore vive.* Paris: Ed. du Seuil, 1975.
12. S. Morard and E. Sanchez, “Conception collaborative d’un jeu d’évasion pédagogique dans le cadre d’une game jam : du design du jeu au design du jouer”, *Sciences du jeu*, no. 16, Oct. 2021, doi: 10.4000/sdj.3517.

Développement et validation d'un assistant pédagogique intelligent pour les personnes enseignantes

Bruno Poellhuber[1], Michel Desmarais[2], Floriane Ronzon[3]

[1] Université de Montréal
Bruno.poellhuber@umontreal.ca
[2] Polytechnique Montréal
michel.desmarais@polymtl.ca
[3] Université de Montréal
floriane.ronzon@umontreal.ca

Résumé. Dans le cadre du projet LAVIA, un assistant pédagogique intelligent a été conçu pour accompagner le nouveau personnel enseignant dans l'analyse de leurs cours, la formulation d'objectifs d'apprentissage et l'élaboration d'un tableau d'alignement pédagogique servant de base à l'élaboration d'un plan de cours. Fondé sur une méthodologie de recherche orientée par la conception et une approche centrée utilisateur, les tests utilisateurs, se fondant sur le modèle TAM, menés auprès de 5 personnes, puis 4 enseignants débutants, avec un protocole de pensée à voix haute, un questionnaire et une entrevue, ont mené à un prototype perçu comme utile pédagogiquement, notamment pour structurer les idées et enrichir les réflexions pédagogiques. Les améliorations apportées à la deuxième version du prototype ont permis d'en accroître la perception d'utilité et de facilité d'utilisation, ce qui nous mène à prévoir un déploiement à moyenne échelle auprès d'un groupe d'enseignants débutants.

Mots-clés : Intelligence artificielle générative, enseignants, plan de cours, alignement pédagogique, planification pédagogique

Abstract. As part of the LAVIA project, an intelligent pedagogical assistant was designed to support new teaching staff in analyzing their courses, formulating learning objectives, and developing a pedagogical alignment table which is the basis of a course suyllabus design. Based on a design-oriented research methodology using a user-centered development approach, user testing guided by the TAM model was conducted with initially five individuals, followed by four novice teachers, employing a think-aloud protocol, a questionnaire, and an interview. This resulted in a prototype perceived as pedagogically useful, particularly for structuring ideas and enriching pedagogical reflections. Improvements made to the second version of the prototype enhanced its perceived usefulness and ease of use, leading us to plan a medium-scale deployment with a group of novice teachers.

Keywords: Generative artificial intelligence, syllabus, teachers, pedagogical alignment, instructional design

1 Introduction et contexte

L'essor récent de l'intelligence artificielle générative (IAg), portée par l'apparition et la mise à disposition au public de grands modèles de langage (LLM) comme chatGPT ou Gemini, modifie en profondeur les pratiques en enseignement supérieur. En enseignement supérieur, 'apparition de ChatGPT a suscité de nombreux débats sur les enjeux liés au plagiat et à l'intégrité académique. L'IAg soulève d'autres préoccupations sur le plan éthique : fiabilité des informations, perte de compétence, manque de transparence, confidentialité des renseignements personnels, propriété intellectuelle, biais, désinformation, etc.

Par ailleurs, l'IAg offre aussi de nombreuses possibilités pour assister les personnes enseignantes (Conseil de l'innovation, 2024; Yan et coll., 2024), notamment dans la préparation de matériel pédagogique, activité en soi moins risquée sur le plan éthique. Par exemple, Karpouzis et coll. (2024) ont développé un agent intelligent qui assiste les personnes enseignantes dans la préparation de leçons.

Les ingénieurs pédagogiques qui accompagnent les personnes enseignantes dans leur processus de planification pédagogique placent la rédaction d'objectifs d'apprentissage et le concept d'alignement pédagogique au cœur de leurs actions. Or les personnes enseignantes éprouvent des difficultés non seulement à formuler adéquatement ces objectifs d'apprentissage mais aussi à les mettre en lien avec les contenus de cours, ainsi que les méthodes d'enseignement et les formules d'évaluation (Filtriani, 2021), ce qui constitue le cœur de l'alignement pédagogique.

Dans le contexte de l'implantation d'un nouveau programme de formation destiné au nouveau personnel enseignant de l'Université de Montréal, plutôt bien disposé à adopter l'IA pour la planification ou la préparation de matériel pédagogique (Lepage, 2024), nous avons entrepris le développement d'un assistant pédagogique intelligent (API), qui vise à les accompagner dans l'analyse de leur cours, la rédaction d'objectifs d'apprentissage et la préparation de leur plan de cours, en nous fondant sur des principes d'éthique par la conception (d'Aquin et al., 2018).

1.1 Objectifs de la recherche

Les objectifs spécifiques de la recherche consistent à : développer de manière itérative un API pour les personnes enseignantes ; évaluer les différentes versions de ces différents prototypes en recourant à des tests utilisateurs, en vue de les améliorer.

2 Cadre conceptuel

2.1 Adoption

L'explication de l'adoption ou non peut quant à elle s'expliquer par plusieurs facteurs qui sont étudiés par des modèles théoriques comme le *TAM - Technology acceptance model* (Davis et al., 1989) selon lequel l'intention d'utiliser une technologie s'explique essentiellement par deux facteurs : la facilité d'utilisation perçue et l'utilité perçue. La

figure 1. montre les éléments retenus des modèles TAM2 et TAM3 (Venkatesh et Bala, 2008).

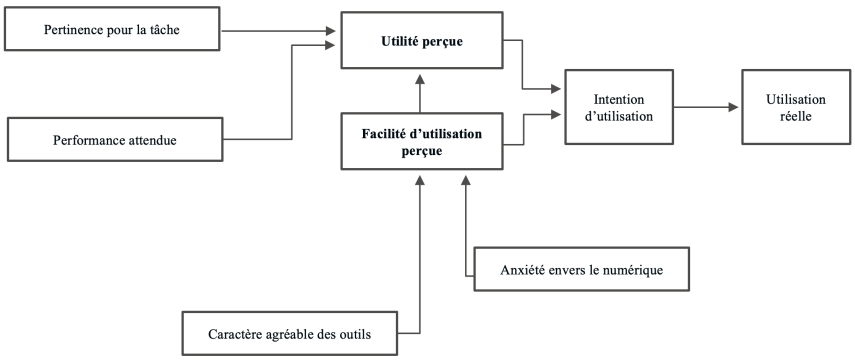


Fig. 1. Échelles mobilisées dans le TAM

3 Méthodologie

Ce projet de recherche recoure à une méthodologie de recherche orientée par la conception ainsi que sur une approche de recherche-développement basée sur le prototypage rapide centré utilisateur. Pour pallier certains des enjeux cités plus haut, nous avons mobilisé une approche d'éthique par la conception où les enjeux éthiques sont considérés dès le début du processus de conception.

Nous appliquons dans ce projet un processus de Conception centrée sur l'utilisateur qui se décline essentiellement en quatre étapes principales : comprendre et spécifier le contexte d'utilisation (ici, celui de l'élaboration d'un plan de cours); comprendre et spécifier les exigences utilisateurs et organisationnelles, produire des solutions de conception (maquettes et prototypes); évaluer les solutions au regard des exigences prédéfinies par le biais de tests utilisateurs.

Les deux premières étapes se sont concrétisées par l'élaboration de scénarios d'utilisation détaillés autour d'une première étape d'analyse du cours (analyse du contexte du cours, des contraintes, des besoins et caractéristiques des personnes apprenantes), aboutissant à la formulation d'objectifs d'apprentissage. La deuxième étape, consiste en l'élaboration d'un tableau d'alignement pédagogique.

Les tests utilisateurs ont été réalisés avec une méthodologie essentiellement qualitative auprès d'un groupe cible de 6 utilisateurs, puis de 4 nouvelles personnes enseignantes et ont misé sur des séances de mise à l'essai d'environ 90 minutes se déroulant en 3 étapes : 1) expérimentation de l'assistant avec un prototocole de pensée à voix haute; 2) un court questionnaire misant sur des questions ouvertes et des échelles tirées du TAM2 et du TAM3, ainsi que sur la perception de valeur affective ou utilitaire; 3) une entrevue faisant un retour sur l'expérimentation mobilisant les thèmes d'utilité, de

facilité d'utilisation et d'expérience utilisateur. Les séances se sont déroulées à distance, sur la plateforme Zoom. Elles ont été enregistrées, puis transcrites et analysées en fonction des étapes suggérées par Miles et Huberman (2003).

4 Résultats

4.1 Premiers tests utilisateurs

Le score d'échelle de chaque participant est présenté dans le tableau 2.

Tableau 1. Tests utilisateurs auprès de 5 enseignants ou CP

TAM Anx Tech	TAM Utilité	TAM Plaisir	TAM Fac Util	TAM Perf Att
Cas 1	2,50	2,50	2,33	3,50
Cas 2	2,75	2,50	2,67	3,75
Cas 3	2,75	3,25	3,67	3,00
Cas 4	2,50	3,75	2,33	3,00

Les retours des participants dans les questions ouvertes du questionnaire ou dans les entrevues ont permis d'identifier plusieurs aspects positifs surtout à l'utilité perçue (voir Tableau 2), ainsi que des difficultés liées à l'interface « J'aurais voulu voir le tableau d'alignement au complet et pouvoir l'éditer directement », liée à la facilité d'utilisation. L'utilité perçue est aussi moindre dans certains contextes, notamment ceux de l'enseignement collégial, ainsi que pour des enseignants expérimentés.

Tableau 2. Utilité perçue de l'API

Utilité perçue +	Commentaires des participants
Analyse des cours	« *Il découpe bien les objectifs d'apprentissage en thématiques logiques.* » (questionnaire) La question sur les défis et les pistes de solution suscite un vif intérêt. (interviewer)
Alignement pédagogique	Le tableau d'alignement semble pertinent pour les enseignants débutants. (interviewer) « *Il m'a permis de verbaliser des attentes pédagogiques avec un vocabulaire plus adapté*. » (Questionnaire)
Soutien à la réflexion pédagogique	La démarche aide à structurer, formuler ou valider des idées pour le plan de cours. (interviewer)

4.2 Deuxième série de tests utilisateurs

Les retours ont mené à des modifications aboutissant à une deuxième version du prototype, ayant été mise à l'essai exclusivement auprès de nouvelles personnes enseignantes à l'université (Tableau 3).

Tableau 3. Tests utilisateurs auprès de 4 enseignants débutants

TAM Anx Tech	TAM Utilité	TAM Plaisir	TAM Fac Util	TAM Perf Att
Cas 1	2,50	5,00	5,00	3,75
Cas 2	2,75	4,00	4,00	3,50
Cas 3	2,50	3,50	3,33	3,50
Cas 4	2,00	5,00	4,33	4,25

Les résultats qualitatifs synthétisés par la chercheure, démontrent différents aspects de l'utilité perçue pour le nouveau personnel enseignant; l'API est *vraiment utile pour les enseignants avec peu d'expérience qui doivent concevoir un cours seuls*, aide à la réflexion à propos du cours, donne des idées, suscite le questionnement, l'ajustement; donne l'impression de discuter avec un collègue. Essentiellement, *les objectifs d'apprentissage proposés par l'assistant ou la rétroaction sur ceux-ci amènent une réflexion sur l'organisation des contenus, sur la formulation et la structure de l'énoncé, amène dimension pédagogique. Il reste un travail à faire.*

5 Discussion et conclusion

Les résultats de la deuxième série de tests indiquent une adéquation marquée entre les besoins des enseignants débutants et les fonctionnalités offertes par l'API. En lien avec le modèle TAM3, les enseignants débutants mettent en évidence plusieurs aspects de l'utilité et une facilité d'utilisation améliorée. L'API est perçue comme un appui à la structuration d'idées, au dialogue pédagogique, à la formulation d'objectifs d'apprentissage clairs et à un alignement pédagogique alimentant la planification de cours. L'API joue également un rôle d'accompagnement réflexif en suscitant des questionnements, en proposant des formulations alternatives et en facilitant la cohérence entre objectifs et activités, via la réflexion pédagogique, ce qui est en adéquation avec les intentions de conception. Ce mécanisme évoque le concept d'échafaudage (Wood, Bruner & Ross, 1976), en soutenant l'autonomie progressive des enseignants dans la conception pédagogique. Sur le plan de la facilité d'utilisation, la lenteur de la réponse et une difficulté de visualisation du tableau d'alignement pédagogique ont été soulevées. Par ailleurs, l'utilité perçue diminue lorsque les plans de cours sont déjà très développés (ex. : au collégial), d'autre part, chez des enseignants très expérimentés. Ces observations rejoignent le concept de renversement de l'expertise **(**Kaluga, 2009**)**, l'utilité perçue étant inversement corrélée au niveau de compétence des utilisateurs.

Du côté des forces, l'assistant pédagogique développé dans le projet LAVIA montre un potentiel prometteur pour soutenir la rédaction d'objectifs d'apprentissage et le processus de planification pédagogique s'incarnant dans un plan de cours, en particulier chez les enseignants débutants. Il facilite la cohérence entre intentions, méthodes et évaluations, tout en stimulant la réflexion pédagogique.

Du côté des limites, la validation de l'API n'a été testée qu'auprès d'un très petit nombre d'utilisateurs, dans une approche essentiellement qualitative ne permettant pas la généralisation.

Dans cette perspective une mise à l'essai élargie auprès dans le cadre du programme de formation des nouveaux enseignants à l'Université de Montréal est prévue pour l'automne 2025. Par ailleurs, l'API, résidant et opérant entièrement dans une infrastructure canadienne AWS, se déclinera aussi avec des fonctions additionnelles dans le futur ; que la conception de plans de leçon, la génération de cas pédagogiques, ou la production de scénarios de simulation, la génération de grilles d'évaluation, etc.

Références

1. Conseil de l'innovation (2024). Prêts pour l'IA. https://conseilinnovation.quebec/wp-content/uploads/2024/02/Rapport_IA_CIQ-1.pdf
2. Yan, L., Sha, L., Zhao, L., Li, Y., Martinez-Maldonado, R., Chen, G., Li, X., Jin, Y. et Gašević, D. (2024). Practical and ethical challenges of large language models in education: A systematic scoping review. *British Journal of Educational Technology*, *55*(1), 90-112. https://bera-journals.onlinelibrary.wiley.com/doi/abs/10.1111/bjet.13370
3. Karpouizis, K., Pantazatos, D., Taouki, J. et Meli, K. (2024, 12 février). *Tailoring Education with GenAI: A New Horizon in Lesson Planning*. arXiv.org. https://arxiv.org/abs/2403.12071v1
4. Fitriani, N. W. (2021). *An Analysis Of Teacher's Lesson Plan For Learning English Through Google Classroom In The Junior High School* (Doctoral dissertation, Universitas Pendidikan Ganesha).
5. Lepage, A. (2024). Étude de l'adoption des principaux types d'usages de l'intelligence artificielle par les enseignants et enseignantes du postsecondaire. Thèse inédite de doctorat. Université de Montréal.
6. d'Aquin, M., Troullinou, P., O'Connor, N. E., Cullen, A., Faller, G., Holden, L. (2018). *Towards an 'Ethics by Design' Methodology for AI Research Projects*. Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society, New York, NY, USA. pp. 54–59. https://doi.org/10.1145/3278721.3278765
7. Davis, F. D. (1989). Technology acceptance model: TAM. *Al-Suqri, MN, Al-Aufi, AS: Information Seeking Behavior and Technology Adoption*, *205*, 219. https://quod.lib.umich.edu/b/busadwp/images/b/1/4/b1409190.0001.001.pdf
8. Venkatesh, V., & Bala, H. (2008). Technology acceptance model 3 and a research agenda on interventions. *Decision sciences*, *39*(2), 273-315.
9. Miles, M. B. et Huberman, A. M. (2003). Analyse des données qualitatives (2e édition). De Boeck Supérieur.
10. Wood, D., Bruner, J. S. et Ross, G. (1976). The role of tutoring in problem solving. *Journal of child psychology and psychiatry*, *17*(2), 89-100.
11. Kalyuga, S. (2009). The expertise reversal effect. In *Managing cognitive load in adaptive multimedia learning* (pp. 58-80). IGI Global.

L'application *MOBILES* : favoriser le partage d'expériences urbaines entre étudiants internationaux

Madjid Sadallah et Marie Lefevre

Universite Claude Bernard Lyon 1, CNRS, INSA Lyon,
LIRIS, UMR5205, 69622, Villeurbanne, France
{madjid.sadallah, marie.lefevre}@liris.cnrs.fr

Résumé. L'intégration des étudiants internationaux, cruciale pour leur réussite, se heurte souvent à des défis majeurs : *isolement social* dû au manque de repères et de réseaux, difficultés de *progression linguistique* limitant les interactions, et besoin d'*engagement actif* dans leur nouvel environnement. Pour tenter d'y répondre, nous présentons *MOBILES*, une application transformant l'exploration urbaine en un outil d'apprentissage et d'intégration. Elle incite les étudiants à explorer activement leur ville, à documenter et partager leurs expériences *via* des annotations géolocalisées, facilitant ainsi leur adaptation socioculturelle et linguistique informelle. Grâce à des recommandations personnalisées basées sur leurs intérêts et activités, *MOBILES* vise à renforcer leur sentiment d'appartenance et à stimuler leur engagement, créant un lien entre intégration et apprentissage situé. Cette démonstration présente les fonctionnalités clés de l'application.

Mots-clé : Étudiants Internationaux · Partage d'Expérience · Exploration Urbaine · Intégration sociale · Apprentissage mobile.

Abstract. The integration of international students, crucial for their success, often faces major challenges: *social isolation* due to a lack of landmarks and networks, difficulties in *linguistic progression* limiting interactions, and the need for *active engagement* in their new environment. To address these needs, we present *MOBILES*, an application transforming urban exploration into a learning and integration tool. It encourages students to actively explore their city, document and share their experiences *via* geolocated annotations, thus facilitating their sociocultural adaptation and informal language practice. Through personalized recommendations based on their interests and activities, *MOBILES* aims to strengthen their sense of belonging and stimulate their engagement, creating a link between integration and situated learning. This demonstration presents the key functionalities of the application.

Keywords: International Students · Experience Sharing · Urban Exploration · Social Integration · Mobile Learning.

1 Introduction

La mobilité internationale étudiante offre des opportunités interculturelles uniques mais confronte souvent les étudiants à des défis majeurs : choc culturel, isolement social et barrières linguistiques [1], affectant leur bien-être et leur réussite. Les dispositifs de soutien classiques (mentorat, cours de langue) sont utiles mais peinent à offrir un accompagnement continu et personnalisé face aux besoins spécifiques d'intégration sociale, de progression linguistique et d'engagement actif [6]. Face à cela, les technologies mobiles géolocalisées offrent un potentiel considérable [3]. Des applications existent pour la navigation, la découverte touristique ou le réseautage social. Cependant, peu d'entre elles combinent spécifiquement le soutien à l'exploration urbaine, la documentation réflexive d'expériences personnelles et le partage au sein d'une communauté d'étudiants internationaux pour répondre de manière intégrée à leurs besoins d'adaptation socio-linguistique et culturelle.

C'est dans ce contexte que nous présentons *MOBILES*, une application développée dans un projet interdisciplinaire (sociolinguistique, géomatique, informatique). *MOBILES* vise à répondre aux besoins identifiés : faciliter l'*intégration sociale* par la découverte de lieux et le partage communautaire ; stimuler la *progression linguistique* par l'incitation à décrire et commenter en contexte ; et favoriser l'*engagement* par une exploration ludique et des recommandations personnalisées. Elle ne remplace pas l'apprentissage formel mais le complète par un soutien informel et situé. Cet article associé à notre démonstration se concentre sur la présentation de l'application et de ses fonctionnalités clés. Les aspects recherche et évaluation font l'objet d'autres publications [5].

2 Le rôle des outils numériques dans l'intégration des étudiants internationaux

Les technologies numériques, telles que les applications mobiles, les systèmes d'information géographique et les plateformes de cartographie participative, sont des outils puissants pour accompagner les étudiants internationaux dans leurs villes d'accueil. Elles favorisent l'apprentissage des langues, la découverte urbaine et le renforcement des liens sociaux et culturels, contribuant ainsi à leur acculturation [2]. Cependant, la plupart des technologies actuelles se concentrent sur des fonctionnalités pratiques telles que la navigation ou l'organisation, et négligent l'intégration sociale et les dimensions émotionnelles et relationnelles, essentielles pour créer un sentiment d'appartenance et renforcer l'engagement des étudiants.

L'intégration des étudiants internationaux dépend de l'enrichissement de leurs expériences, essentielles à leur adaptation et épanouissement. Il est crucial de développer des outils qui non seulement améliorent ces expériences, mais aussi les documentent, les partagent, les analysent et encouragent une introspection constructive. Ces fonctionnalités aident les étudiants à comprendre leur parcours, à y trouver du sens et à interagir plus profondément avec leur environnement. L'analyse de ces expériences permet de concevoir des outils mieux adaptés qui,

en se basant sur leurs préférences et interactions passées, peuvent guider les étudiants vers des lieux et activités pertinents. En combinant navigation, planification et interactions culturelles, ces outils réduisent l'isolement, renforcent l'inclusion et encouragent un engagement actif, tout en offrant des moyens de documenter, réfléchir et partager leurs expériences pour une intégration plus harmonieuse.

3 L'application *MOBILES*

L'application *MOBILES*, développée de manière participative dans le cadre d'un projet pluridisciplinaire[1], vise à accompagner les étudiants internationaux dans leur découverte de leur ville d'étude.

Elle leur permet de documenter leurs expériences et de partager leurs perspectives. Son architecture modulaire intègre plusieurs composants interconnectés pour favoriser une exploration enrichissante et une documentation immersive. Les détails techniques de mise en œuvre, disponibles dans [4], seront illustrés lors de la démonstration.



(a) Page d'accueil (b) Une annotation (c) Un parcours

Fig. 1: Éléments de l'interface de *MOBILES*.

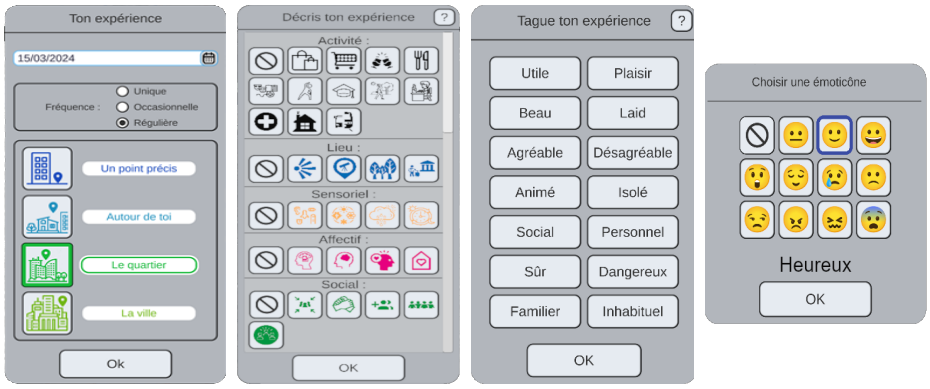
[1] https://mobiles-projet.huma-num.fr/

3.1 Interface principale

L'interface s'articule autour d'une vue cartographique interactive permettant la localisation, la recherche de points d'intérêt et la visualisation des expériences partagées (annotations, parcours) par la communauté (Figure 1). Cela crée une plateforme riche de partage d'expériences, soutenant ainsi la base de connaissances de la communauté étudiante et contribuant à l'intégration sociale en rendant visible l'activité des pairs.

3.2 Création et partage d'annotation

Le processus de création d'une annotation (Figure 2) débute par la saisie des informations temporelles. La date de l'expérience est automatiquement enregistrée mais peut être modifiée manuellement pour permettre l'annotation rétrospective. L'étudiant renseigne également sa perception de la fréquence (*unique*, *occasionnelle*, *régulière*) ; cette donnée subjective, distincte d'une mesure objective de visite, sert d'indicateur qualitatif de l'intérêt perçu pour le lieu, information utile aux recommandations. Il définit ensuite l'étendue spatiale (*point spécifique*, *autour*, *quartier*, *ville*).



(a) Contexte (b) Liste d'icônes (c) Liste de tags (d) Emoticônes

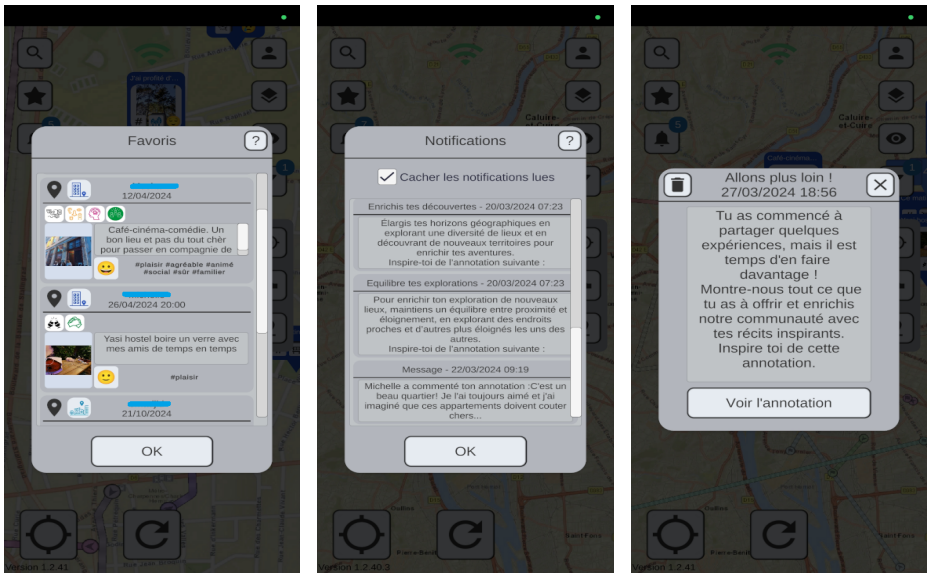
Fig. 2: Composants essentiels du processus de création d'annotation.

L'annotation est ensuite constituée d'un texte et/ou d'une photo, et peut être enrichie par des icônes, balises textuelles (tags) et émoticônes. Ce processus de documentation structurée vise plusieurs objectifs : il soutient l'engagement et la réflexivité de l'étudiant ; il offre un cadre pour la progression linguistique par la description contextualisée ; enfin, le partage de ces informations détaillées facilite l'intégration sociale par la mise en commun d'expériences.

Les niveaux de visibilité (publiques, groupes, privées) permettent un contrôle fin du partage.

3.3 Création et partage de parcours

Les étudiants peuvent tracer leurs parcours urbains (Figure 1c), enregistrant trajets et points d'intérêt. Chaque point séquentiel peut être annoté. Ces parcours structurent l'exploration (renforçant l'engagement et la motivation) et permettent de partager des itinéraires thématiques (favorisant l'intégration sociale et la découverte). L'application propose trois options de création de parcours : (1) *Suivi GPS automatique* pour enregistrer en temps réel ; (2) *Cartographie manuelle* pour tracer directement dans l'application ; (3) *Séquençage rétrospectif* pour connecter les annotations en une séquence significative. Comme pour les annotations, l'application offre les mêmes niveaux de visibilité pour gérer le partage des parcours.



(a) Gestionnaire de favoris (b) Centre de notification (c) Une recommandation

Fig. 3: Quelques fonctionnalités de l'application *MOBILES*.

3.4 Exploration des annotations et parcours partagés

Les étudiants peuvent explorer les annotations et les parcours sur la carte interactive de l'application, qui offre des filtres pour visualiser des catégories spécifiques en fonction de la localisation, de la date, des icônes, des tags ou des émotions associées. Ces filtres permettent de découvrir une grande variété d'expériences urbaines, d'identifier des zones d'intérêt et de partager des perceptions sur des lieux communs. L'application facilite également les interactions sociales en permettant

aux étudiants de commenter et d'interagir avec les annotations et les parcours des autres, favorisant ainsi un espace d'échange dynamique et renforçant l'aspect communautaire de la plateforme. Enfin, le *Gestionnaire de favoris* (Figure 3a) permet de sauvegarder des annotations, des trajets et des points d'intérêt pour des visites futures.

3.5 Système de recommandations personnalisées

L'application enregistre toutes les données générées (annotations, parcours, commentaires, etc.) et collecte l'ensemble des interactions des étudiants. En s'appuyant sur l'analyse de ces données, *MOBILES* intègre un système de recommandations personnalisées (Figure 3c). Ce système emploie une approche hybride, considérant notamment les interactions passées de l'utilisateur, la similarité de son profil avec d'autres utilisateurs, et la popularité des contenus au sein de la communauté. Il a pour double objectif de suggérer de nouvelles explorations pertinentes (visant l'engagement et l'intégration sociale) et d'inciter à enrichir les annotations produites (soutenant la progression linguistique et la réflexivité). La présentation détaillée et l'évaluation de ce système sont présentées dans [5]. Un module de notifications (Figure 3b) tient les utilisateurs informés des nouvelles recommandations, interactions et publications, facilitant ainsi le suivi de leurs activités.

4 Conclusion

L'application *MOBILES*, conçue pour faciliter l'intégration des étudiants internationaux, transforme l'exploration urbaine en un dispositif d'apprentissage personnalisé. Par une approche fondée sur la découverte active et le partage d'expériences, elle favorise leur compréhension du nouvel environnement et leur adaptation socioculturelle. Les recommandations ciblées de *MOBILES*, dont l'efficacité est suggérée par des résultats d'une étude préliminaire à Lyon [5], contribuent à une intégration sociale et académique réussie. Enfin, la conception de l'interface, notamment l'usage des éléments iconographiques, fait l'objet d'une attention particulière ; des tests d'utilisabilité sont prévus pour optimiser l'équilibre entre richesse fonctionnelle et clarté visuelle. Ces observations, bien que nécessitant des investigations complémentaires, soulignent le potentiel des technologies numériques pour améliorer substantiellement l'expérience des étudiants internationaux, faciliter leur transition et enrichir leur parcours global.

Remerciements

Ce travail est soutenu par le projet MOBILES (ANR-20-CE38-0009) financé par l'Agence Nationale de la Recherche (ANR).

References

1. Gutema, D.M., Pant, S., Nikou, S.: Exploring key themes and trends in international student mobility research—a systematic literature review. Journal of Applied Research in Higher Education **16**(3), 843–861 (2024)
2. Hofhuis, J., van Egmond, M.C., Lutz, F.E., von Reventlow, K., Rohmann, A.: The effect of social network sites on international students' acculturation, adaptation, and wellbeing. Frontiers in Communication **8**, 1186527 (2023)
3. Kukulska-Hulme, A.: Mobile and personal learning for newcomers to a city. Electronic Journal of Foreign Language Teaching **17**(1), 93–103 (2020)
4. Lefevre, M., Benedetto, R., Sadallah, M., Chevalier, D., Cunty, C., Grassin, J.F., Joliveau, T., Mathian, H., Scheffler, C.: MOBILES: A Mobile Application for Documenting the Urban and Social Experiences of International Students. Software, Version 1.2.41 (2024), https://hal.science/hal-04771100
5. Sadallah, M., Lefevre, M.: Personalized recommender system for improving urban exploration and experience documentation of international students. In: Proceedings of the 17th International Conference on Computer Supported Education (CSEDU 2025) - Volume 2. pp. 923–930. SCITEPRESS - Science and Technology Publications (2025). https://doi.org/10.5220/0013202800003932
6. Tavares, V.: Feeling excluded: International students experience equity, diversity and inclusion. International Journal of Inclusive Education **28**(8), 1551–1568 (2024)

Co-conception d'un jeu collaboratif multimodal pour le développement des compétences spatiales chez les élèves du primaire

Alain Sénécail[1], Mireille Bétrancourt[1], Gaëlle Molinari[1], Éric Sanchez[1], Fatou-Maty Diouf[1]

[1] TECFA – FPSE, Université de Genève

Résumé. Le projet Deep Space a pour objectif la conception d'un EIAH visant à développer les compétences spatiales des élèves à l'école primaire. Une technologie particulièrement pertinente est la réalité virtuelle (sur ordinateur ou immersive), parce qu'elle permet de représenter des espaces à large échelle et possède des potentiels pédagogiques, technologiques et sociaux bien documentés. Pour autant, les recherches avec des élèves d'école primaire sont rares. L'approche pédagogique choisie est un format collaboratif basé sur le jeu, dont l'efficacité pour engager les élèves dans le raisonnement et la résolution de problèmes a été démontrée. Cependant, ces environnements sont exigeants en termes d'orchestration par les enseignants, notamment pour garantir une participation équitable des élèves. Cet article décrit les fondements théoriques sur lesquels s'appuie la conception du dispositif mixte (numérique et physique) et la méthode de recherche orientée conception mise en œuvre.

Mots-clés : cognition spatiale, réalité virtuelle, apprentissage collaboratif, jeu, recherche orientée par la conception

Abstract. The Deep Space project aims to create a Technology-Enhanced Learning Environment (TELE) to help primary school students develop their spatial skills. Virtual reality (VR), whether computer-based or immersive, is particularly useful because it allows students to explore large spaces and offers many benefits for learning and interaction. Yet, studies involving primary school students remain rare. The project uses a collaborative, game-based approach, which has proven effective in helping students think and solve problems. These tools, however, can be challenging for teachers to manage, especially to ensure all students participate equally. This article explains the ideas behind the design of the mixed (digital and physical) system and the research method used to develop it.

Keywords: spatial cognition, virtual reality, collaborative learning, game-based learning, design-based research

1 Introduction

La cognition spatiale, définie comme la capacité à se représenter et raisonner sur l'espace [1], est corrélée à la réussite scolaire dans des domaines comme les sciences, les mathématiques, la géographie, l'ingénierie [2]. Elle joue également un rôle central dans la résolution de problèmes complexes, le raisonnement inférentiel et la compréhension abstraite [3]. Pourtant, les recherches soulignent des inégalités précoces d'acquisition, ce qui peut impacter les trajectoires scolaires des élèves [4]. Cependant, peu d'études portent sur l'entraînement des compétences spatiales à l'école primaire dans l'objectif de réduire ces inégalités et de favoriser leur transfert dans les tâches scolaires [2]. C'est pourquoi cette recherche s'intéresse au développement, chez les élèves âgés entre 7 et 10 ans, de compétences en lien avec le macro-espace, comme la navigation, le repérage et la prise de perspectives visuo-spatiale.

Dans ce contexte, le projet DEEP Space vise à concevoir un dispositif multimodal (numérique et physique) pour soutenir le développement équitable de ces compétences spatiales chez les élèves. La réalité virtuelle (qu'elle soit sur ordinateur ou immersive) s'avère ici particulièrement pertinente, car elle permet de représenter des espaces à large échelle. Ses potentiels pédagogiques, technologiques et sociaux sont en outre bien documentés [5, 6].

Plus spécifiquement, un dispositif basé sur le jeu, collaboratif et multimodal, soutenu par des technologies immersives, pourrait constituer un levier pour favoriser l'apprentissage des compétences spatiales tout en contribuant à réduire les inégalités d'accès et d'engagement. Toutefois, de tels environnements pédagogiques posent des défis notables en matière d'orchestration : mobiliser les élèves, maintenir une dynamique collaborative et assurer une participation équilibrée ne vont pas de soi [7].

Le projet s'articule ainsi autour de quatre dimensions clés : les propriétés de l'environnement virtuel, la dynamique de collaboration entre élèves, les stratégies d'orchestration déployées par les enseignants, et les compétences spatiales ciblées.

Quelles sont alors les caractéristiques d'un environnement d'apprentissage multimodal et collaboratif — combinant réalité virtuelle, supports physiques et mécaniques de jeu — permettant de développer de manière équitable les compétences spatiales liées au macro-espace chez les élèves du primaire, et quelle méthodologie adopter pour le concevoir et l'orchestrer efficacement en contexte scolaire ?

2 Cognition spatiale, réalité virtuelle et jeu collaboratif

2.1 Réalité virtuelle et cognition spatiale

La réalité virtuelle (RV), qu'elle soit immersive ou non, permet la simulation d'environnements à grande échelle, offrant des situations d'apprentissage engageantes qui mobilisent les compétences de repérage, d'orientation et de navigation [5]. Ces environnements sollicitent la co-construction de cartes cognitives, encodées selon plusieurs

cadres de référence utilisés en parallèle (égocentré, allocentré, intrinsèque) [8]. La représentation ne se limite pas au mental : elle s'instancie aussi à travers des registres verbaux (ex. : vocabulaire spatial) et graphiques (ex. : cartes), abordés à l'école primaire mais pas nécessairement acquise.

Des études montrent que la VR peut renforcer les apprentissages spatiaux dans les disciplines STEM [9], en particulier chez les apprenants avec des compétences spatiales faibles [10]. Toutefois, d'autres recherches révèlent des effets mitigés voire négatifs sur les performances, selon le design de l'activité ou la population ciblée [11]. Peu de travaux concernent les élèves de primaire, à l'exception du projet SPAGEO qui a permis d'élaborer et tester des séquences didactiques utilisant la VR pour des enfants de 7 à 10 ans, avec des effets encourageants [12].

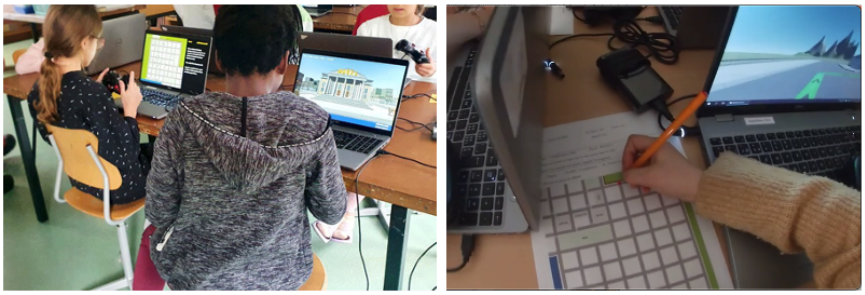


Fig.1. Mise en œuvre d'une séquence didactique dans le cadre du projet Spageo [12]

2.2 Jeu collaboratif et cognition spatiale

Les jeux éducatifs numériques allient motivation intrinsèque et espace virtuel propice à l'apprentissage, la pratique, l'erreur et le développement. De nombreuses études soulignent l'intérêt de ces jeux éducatifs, notamment en enseignement primaire [10, 13] et dans des contextes collaboratifs [14]. L'apprentissage collaboratif consiste à engager les élèves dans un effort conjoint de partage d'informations, de coordination et de collaboration pour résoudre une tâche commune. L'étude menée par Tversky et Hard [15] montre qu'en travaillant ensemble, les enfants sont plus enclins à adopter les perspectives de leurs pairs, en considérant les relations spatiales non seulement à partir de leurs propres expériences, mais aussi de celles des autres. Cette prise de perspective, un élément clé de l'interaction sociale efficace, s'étend au-delà des actions physiques à la compréhension et à l'interprétation de l'environnement. Viana et al. [16] démontrent que les enfants réussissent mieux les tâches de transformation spatiale lorsqu'ils les résolvent avec un partenaire plutôt que seuls. Plus précisément, c'est la théorie de l'esprit que ce projet vise à entrainer. En effet, définie comme la capacité à attribuer des états et des processus mentaux et émotionnels à soi-même et aux autres, était un meilleur prédicteur de la performance spatiale des élèves en conditions dyadiques que leur âge, leur sexe et leur performance spatiale en situations individuelles [16].

2.3 Équité, genre et orchestration pédagogique

Les inégalités de performance en cognition spatiale, souvent observées dès le jeune âge, touchent en particulier les filles, qui sont statistiquement moins exposées à certaines activités spatiales dans leurs environnements familiaux ou scolaires [4]. Ces écarts ne sont ni biologiquement déterminés, ni inévitables, mais largement liés à des facteurs sociaux, culturels et pédagogiques. Dans cette perspective, il semble essentiel de concevoir des environnements d'apprentissage qui soient explicitement pensés pour réduire ces inégalités. Cela implique d'intégrer, dès la phase de conception, des mécanismes d'équité structurelle [17], tels que :

- des rôles favorisant une participation équilibrée ;
- des tâches interdépendantes valorisant des compétences variées ;
- et des dispositifs de reconnaissance des contributions, visibles et explicites, pour renforcer l'engagement de tous les profils d'élèves.

En parallèle, ces environnements d'apprentissage collaboratifs exigent une orchestration pédagogique fine : les enseignants doivent pouvoir gérer les transitions entre supports, réguler les interactions au sein des groupes, et mobiliser les ressources au bon moment selon les objectifs visés [7]. L'orchestration ne se réduit pas à un ajustement en aval : elle doit être intégrée dès les premières étapes du design, en tenant compte de la planification pédagogique, de l'animation en classe et de la participation des enseignants à la co-conception, en lien avec les choix technologiques et didactiques.

3 Méthode : recherche orientée par la conception

Le projet adopte une approche de recherche orientée par la conception, inscrite dans une perspective pragmatique, collaborative et itérative. Cette démarche associe enseignants du primaire, chercheurs en sciences de l'éducation et en technologies éducatives, ainsi qu'une développeuse. Les enseignants ne sont pas considérés comme des sujets d'étude, mais bien comme des partenaires à part entière du processus de conception et de la recherche.

Le processus repose sur un cycle itératif structuré à deux niveaux interdépendants :

- un premier niveau centré sur la co-conception du jeu : il comprend l'identification des besoins, la définition des compétences spatiales visées, l'élaboration des scénarios, et les premiers choix technico-pédagogiques ;
- un deuxième niveau, en continuité, qui permet l'itération et la mise à l'épreuve des conjectures et prototypes dans les contextes réels d'usage. C'est à ce niveau que s'affinent les questions de recherche, les indicateurs d'évaluation et les analyses empiriques.

Ce modèle n'est pas séquentiel : les dimensions théoriques, méthodologiques et empiriques interagissent dès les premières phases. Les conjectures et objectifs évoluent en fonction des échanges et expérimentations.

Les ateliers de co-design sont donc envisagés comme des moments fondateurs du processus. Ils permettent de reconfigurer les problématiques, de faire émerger des pistes de conception réalistes, et de mettre à l'épreuve les modèles théoriques en situation.

Cette approche systémique vise ainsi à renforcer les allers-retours entre recherche et pratique, en tenant compte des contraintes institutionnelles, pédagogiques et techniques.

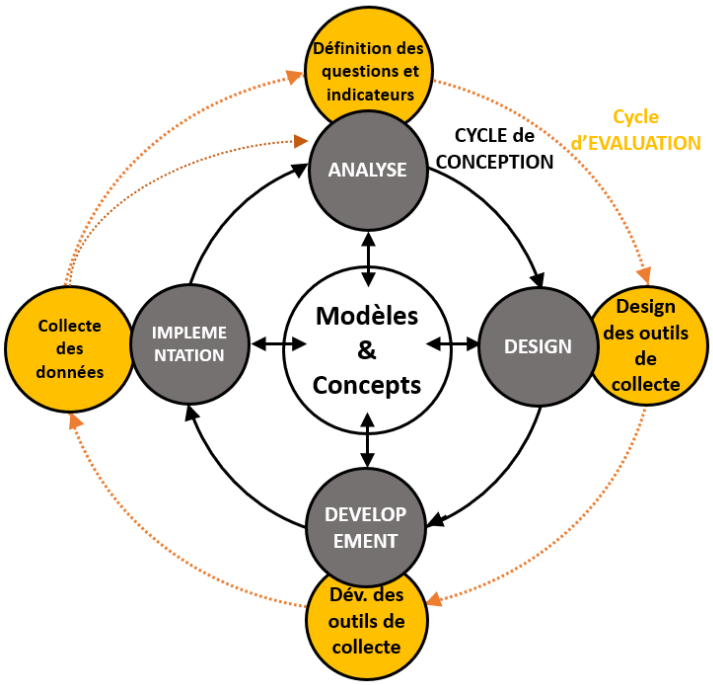


Fig.2. Le cycle de recherche orientée par la conception, adopté dans le projet (adapté de [18])

4 Prototype initial et perspectives

Deux premières journées de co-design, sous forme de Game Jam, ont réuni six enseignants du primaire, un développeur et des chercheurs en éducation (Fig. 3). Cette collaboration a permis de poser les fondements d'un prototype de jeu coopératif basé sur une aventure narrative : sauver une ville médiévale d'une malédiction en explorant ses lieux à la recherche d'artéfacts.



Fig.3. Séances de co-conception entre enseignants, chercheurs et développeur

Le jeu mobilise trois supports complémentaires :

- un plateau physique avec cartes, repères et objets tangibles (Fig. 4) ;
- une tablette numérique donnant accès à des cartes, indices et fiches de mission ;
- un casque de réalité virtuelle permettant l'exploration immersive de certains lieux.



Fig.4. Prototype du plateau de jeu

Par petits groupes, les joueurs explorent, coopèrent et résolvent des énigmes spatiales, en construisant une carte interactive évolutive, nourrie d'indices physiques (cartes, photos) et d'environnements numériques (plans, vues immersives). Ce processus soutient le passage d'un cadre égocentré (local, subjectif) à un cadre allocentré (global, partagé) de l'espace.

Le gameplay repose sur trois mécaniques principales :

- exploration multimodale : alternance entre immersion, cartes et supports textuels, stimulant la prise de perspective ;
- collaboration interdépendante : chaque joueur dispose d'informations partielles à partager ;
- validation dynamique : progression basée sur des feedbacks en temps réel et un système de points valorisant les stratégies collectives.

Ce prototype tente ainsi de répondre à une problématique complexe : comment concevoir un environnement hybride et orchestrable, capable de développer équitablement les compétences spatiales liées au macro-espace en contexte scolaire.

Cette première phase de co-design a permis de faire émerger des mécaniques de jeu cohérentes avec les objectifs d'apprentissage et les contraintes d'orchestration, illustrant l'intérêt d'une co-conception étroite avec les enseignants. Toutefois, la nature itérative de la démarche DBR, bien qu'indispensable, s'est révélée exigeante en termes de coordination, de temporalité et d'engagement des acteurs. Ces tensions invitent alors à repenser certains formats de collaboration pour en assurer la pérennité. Elles confirment aussi la pertinence d'une approche DBR, qui articule ajustements progressifs du prototype et production de connaissances transférables. Les prochaines phases viseront à stabiliser le dispositif, à tester sa robustesse en contexte réel et à analyser ses effets sur les apprentissages spatiaux et les pratiques enseignantes.

Références

1. Tversky, B. (2005). Visuospatial reasoning. In K. J. Holyoak & R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 209–249). Cambridge University Press.
2. Lowrie, T., & Jorgensen, R. (2018). Equity and spatial reasoning: Reducing the mathematical achievement gap in gender and social disadvantage. *Mathematics Education Research, 30*, 65–75. https://doi.org/10.1007/s13394-017-0213-7
3. Delahaye, M., Lemoine, P., Cartwright, S., et al. (2015). Learning aptitude, spatial orientation and cognitive flexibility tested in a virtual labyrinth after virtual stress induction. *BMC Psychology, 3*(22), 1–11.
4. Levine, S. C., Foley, A., Lourenco, S., Ehrlich, S., & Ratliff, K. (2016). Sex differences in spatial cognition: Advancing the conversation. *Wiley Interdisciplinary Reviews: Cognitive Science, 7*(2), 127–155. https://doi.org/10.1002/wcs.1380
5. Makransky, G., & Petersen, G. B. (2021). The cognitive affective model of immersive learning (CAMIL): A theoretical research-based model of learning in immersive virtual reality. *Educational Psychology Review, 1–22.*
6. Makransky, G., & Petersen, G. B. (2023). The theory of immersive collaborative learning (TICOL). *Educational Psychology Review, 35*(4), 103.
7. Olsen, J. K., Rummel, N., & Aleven, V. (2021). Designing for the co-orchestration of social transitions between individual, small-group, and whole-class learning in the classroom. *International Journal of Artificial Intelligence in Education, 31*, 24–56. https://doi.org/10.1007/s40593-020-00228-w
8. Tamborello, F. P., Sun, Y., & Wang, H. (2012). Spatial reasoning with multiple intrinsic frames of reference. *Experimental Psychology, 59*(3).
9. Clark, D. B., Tanner-Smith, E. E., & Killingsworth, S. S. (2016). Digital games, design, and learning: A systematic review and meta-analysis. *Review of Educational Research, 86*(1), 79–122.
10. Newcombe, N. S., Hegarty, M., & Uttal, D. (2023). Building a cognitive science of human variation: Individual differences in spatial navigation. *Topics in Cognitive Science, 15*(1), 6–14. https://doi.org/10.1111/tops.12626
11. Parong, J. (2021). Multimedia learning in virtual and mixed reality. In R. E. Mayer & L. Fiorella (Eds.), *The Cambridge Handbook of Multimedia Learning* (3rd ed., pp. 498–509). Cambridge University Press. http://doi.org/10.1017/9781108894333.051
12. Bénard-Linh Quang, S., Berney, S., Matri, S., Coutat-Gousseau, S., & Diouf, F.-M. (2024). Les enjeux des connaissances spatiales pour la conception d'un environnement virtuel pour l'enseignement de la géométrie au primaire. In N. Duroisin & N. Loye (Eds.), *Évaluation*

des apprentissages et technologies numériques : évolution, nouveautés et défis actuels (pp. 59–100). Peter Lang. https://doi.org/10.3726/b21721

13. Sailer, M., & Homner, L. (2020). The gamification of learning: A meta-analysis. *Educational Psychology Review, 32*(1), 77–112.
14. Jeong, H., Hmelo-Silver, C. E., & Jo, K. (2019). Ten years of computer-supported collaborative learning: A meta-analysis of CSCL in STEM education during 2005–2014. Educational research review, 28, 100284.
15. Tversky, B., & Hard, B. M. (2009). Embodied and disembodied cognition: Spatial perspective-taking. *Cognition, 110*(1), 124–129.
16. Viana, K. M., Zambrana, I. M., Karevold, E. B., & Pons, F. (2016). Beyond conceptual knowledge: The impact of children's theory of mind on dyadic spatial tasks. *Frontiers in Psychology, 7*, 1635.
17. Johnson, D. W., & Johnson, R. T. (2009). An educational psychology success story: Social interdependence theory and cooperative learning. Educational researcher, 38(5), 365-379.
18. Prior, E., Mandran, N., Monnier, S., Jaccard, D. (2025). co.LAB: A Web Platform Dedicated to the Design and Evaluation of Learning Games. In: Leahy, M., Reffay, C. (eds) Digitally Transformed Education: Are We There Yet?. OCCE 2024. IFIP Advances in Information and Communication Technology, vol 734. Springer, Cham.

Modèle de recommandation stochastique et de raisonnement à partir de cas appliqué à l'EIAH AI-VT

Daniel Soto-Forero[1 [0000−0003−0753−4673]], Marie-Laure Betbeder[1 [0000−0002−8103−4098]], and Julien Henriet[1 [0000−0002−7671−4574]]

DISC, Université Marie et Louis Pasteur, CNRS, Institut FEMTO-ST 16 Route de Gray, 25000, Besançon, France

Résumé Cet article présente l'intégration d'un modèle de recommandation fondé sur l'échantillonnage de Thompson (ET) à un modèle multi-agent associé au raisonnement à partir de cas (ESCBR-SMA). Le modèle intégré est utilisé pour améliorer l'adaptation en temps réel d'un environnement informatique pour l'apprentissage humain (EIAH) appelé AI-VT. Il a été évalué à partir de deux bases de données d'apprenants afin de prédire les notes d'exercices de différents niveaux de complexité selon trois scénarios prédéfinis. Nous avons comparé l'algorithme de recommandation et l'intégration de l'algorithme de recommandation avec l'algorithme de prediction en évaluant l'évolution de l'acquisition des connaissances de chaque apprenant. Les mesures effectuées permettent de déterminer la précision de la prédiction pour chaque apprenant et chaque niveau de complexité. Les résultats montrent que l'intégration de l'adaptation stochastique et de la prédiction avec le paradigme du raisonnement à partir de cas permet d'obtenir une meilleure identification des faiblesses des apprenants et donc une meilleure recommandation de parcours personnalisé pour l'apprenant.

Keywords: Environnement Informatique d'Apprentissage Humain · Raisonnement à Partir de Cas · Stacking · Régression · Recommandation Stochastique

Abstract. This paper presents the integration of a recommendation model based on Thompson sampling (TS) with a multi-agent model associated with case-based reasoning (ESCBR-SMA). The integrated model is used to improve the real-time adaptation of an intelligent tutoring system (ITS) called AI-VT. It was evaluated on two learner databases to predict scores for exercises of different levels of complexity under three predefined scenarios. We compared the recommendation algorithm and the integration of the recommendation algorithm with the prediction algorithm by evaluating the evolution of each learner's knowledge acquisition. The measurements carried out allow us to determine the accuracy of the prediction for each learner and each level of complexity. The results show that integrating stochastic adaptation and prediction with the

case-based reasoning paradigm results in better identification of learner weaknesses, and therefore in better personalized path recommendations for the learner.

Keywords: Intelligent Tutoring System · Case-Based Reasoning · Stacking · Regression · Stochastic Recommander

1 Introduction

Le système AI-VT (Artificial Intelligence - Virtual Trainer) est un environnement informatique pour l'apprentissage humain (EIAH) créé pour aider les apprenants à comprendre et à acquérir des connaissances dans divers domaines [10,9].

Certains EIAH intègrent un module de recommandation permettant de personnaliser les contenus et les exercices en fonction des besoins et des résultats des apprenants. L'efficacité du système dans l'acquisition des connaissances et l'adaptation aux différents types d'apprentissage dépend de ce module [6].

Les contributions de cet article sont les suivantes :

- proposition d'un modèle de recommandation (échantillonnage de Thompson - ET) intégré à un système de raisonnement à partir de cas (RàPC) couplé à un système multi-agents (ESCBR-SMA),
- évaluation du modèle intégré pour l'amélioration de la personnalisation du parcours de l'apprenant utilisant l'EIAH AI-VT.

La section 2 de cet article présente les travaux connexes aux EIAH, le RàPC, les techniques d'apprentissage d'ensemble (*Ensemble Stacking*), ET et la régression. Le modèle proposé est expliqué dans la section 3. La section 4 décrit le protocole expérimental utilisé, présente et discute les résultats. Enfin, les conclusions et les travaux futurs sont abordés dans la section 5.

2 Travaux Connexes

Pour améliorer l'expérience de l'apprenant, certains EIAH adaptent le contenu aux besoins et aux connaissances des apprenants afin de leur permettre de progresser dans l'acquisition de connaissances [1] [11].

Mao *et al.* [7] proposent un système de recommandation bilatéral fondé sur l'apprentissage par renforcement (RTR) pour personnaliser des quiz en choisissant des questions adaptées au niveau des apprenants.

Les informations historiques et les connaissances acquises par l'application complètent l'utilisation des algorithmes d'intelligence artificielle en améliorant l'adaptation du système, comme l'illustre la méthode de recommandation de Li *et al.* [5]. Testée sur les bases de données *Assistment 2009-201 et Eedi 2020*, cette méthode peut améliorer les performances des apprenants en recommandant des exercices appropriés. Le travail de Clemente *et al.* [3] infère le progrès des

apprenants grâce à une ontologie dynamique pour détecter, corriger les faiblesses et adapter le système de recommandation.

Clément *et al.* [2] font une analogie entre les EIAH et le problème du *Bandit Manchot* en se fondant sur le fait que dans les deux cas, il est nécessaire de trouver l'ensemble d'actions maximisant le gain. Cela se résume à ce que l'on appelle le compromis "*exploration/exploitation*" dans l'apprentissage automatique. Le modèle proposé permet à l'enseignant de définir le parcours du cours en fonction des réponses des élèves.

Le travail de Soto *et al.* [8], utilise un algorithme fondé sur l'ET, la recommandation dans les EIAH étant en effet à la fois différente pour chaque apprenant et dynamique dans le temps.

3 Modèle Proposé

Le modèle proposé est une intégration d'un modèle de recommandation fondé sur l'échantillonnage de Thompson (ET) à un modèle multi-agent associé au raisonnement à partir de cas (ESCBR-SMA). Le modèle de recommandation produit une adaptation en fonction des notes de l'apprenant et l'ESCBR-SMA effectue une prédiction pour valider l'adaptation générée.

Une fonction de pondération unifie les résultats du modèle. La recommandation finale est celle qui optimise l'expression 3.

Ce modèle adapte grâce à l'ET, puis l'ECBR-SMA propose une prédiction et enfin le processus se termina par la prise de décision envoyée à l'apprenant. Le tableau 1 présente les variables et les paramètres du modèle proposé ainsi que les mesures employées.

Table 1. Paramètres (p), variables (v) et fonctions (f) du modèle proposé

ID	Type	Description	Domain
α	p	Paramètre de la distribution BETA	$[1, \infty] \in \mathbb{R}$
β	p	Paramètre de la distribution BETA	$[1, \infty] \in \mathbb{R}$
t	p	Temps défini des itérations	$\mathbb{N}$
c	p	Niveau de complexité	$\mathbb{N}$
$k_{t,c}$	v	Évolution de la connaissance dans le temps t pour le niveau de complexité c	$[0, 1] \in \mathbb{R}$
$vk_{t,c}$	v	Évolution de la connaissance pour chaque niveau de complexité c	$\mathbb{R}$
TS_c	v	Récompense de l'ET pour un niveau de complexité c	$[0, 1] \in \mathbb{R}$
TSN_c	v	Normalisation de TS_c avec d'autres niveaux de complexité	$[0, 1] \in \mathbb{R}$
$ESCBR_c$	v	Prédiction de la note pour un niveau de complexité c	$\mathbb{R}_+$
D_{JS}	f	Divergence de Jensen-Shannon	$[0, 1] \in \mathbb{R}$

L'intégration se fait en trois étapes. Tout d'abord, il est nécessaire d'avoir des valeurs aléatoires pour chaque niveau de complexité c en utilisant les distribu-

tions de probabilités générées avec le modèle ET (équation 1). La normalisation de toutes ces valeurs est calculée selon l'équation 2. Ces valeurs servent de paramètres de priorité pour les prédictions effectuées par le modèle ESCBR-SMA dans l'équation 3.

$$TS_c = rand(Beta(\alpha_c, \beta_c)) \tag{1}$$

$$TSN_c = \frac{TS_c}{\sum_{i=0}^{4} TS_i} \tag{2}$$

$$n_c = max(TSN_c * ESCBR_c) \tag{3}$$

Après avoir calculé les valeurs finales pour chaque niveau de complexité, le système propose le niveau de complexité ayant la valeur la plus élevée (équation 3).

4 Résultats et Discussion

Deux bases de données ont été utilisées pour tester le modèle, la première a été générée [8] et la seconde est un ensemble de données réelles d'étudiants [4].

La première série de tests est définie sous la forme de différents scénarios. Dans le scénario E1, la note d'un apprenant au premier niveau de complexité est interpolée après 3 questions. Dans le scénario E2, la note de la 9ème question dans le même niveau de complexité est interpolée. Le scénario E3 interpole les deux notes dans un niveau de complexité supérieur.

Le modèle a été comparé à neuf algorithmes de régression présentés dans le tableau 2.

Table 2. Liste des algorithmes évalués

ID	Algorithme	ID	Algorithme
A1	Linear Regression	A6	Polinomial Regression
A2	K-Nearest Neighbor	A7	Ridge Regression
A3	Decision Tree	A8	Lasso Regression
A4	Random Forest (Ensemble)	A9	Gradient Boosting (Ensemble)
A5	Multi Layer Perceptron	A10	Proposed ET-ESCBR-SMA

Les algorithmes ont été évalués à l'aide de trois mesures (Root Mean Squared Error - RMSE, Median Absolute Error - MedAE, Mean Absolute Error - MAE). Les résultats figurent dans le tableau 3. L'algorithme proposé obtient de meilleurs résultats que les autres algorithmes avec lesquels il a été comparé dans quatre des neuf scénarios. Il se place en deuxième position dans les cinq autres scénarios.

La progression dans le temps avec ET est évaluée avec la divergence de Jensen-Shannon (équations 4 et 5) et la famille de distribution Beta en t et $t-1$ comme second test.

Table 3. Résultats de régression avec 100 exécutions

	Algorithme									
Scenario (Métrique)	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10
E1 (RMSE)	0.625	0.565	0.741	0.56	0.606	0.626	0.626	0.681	0.541	**0.54**
E1 (MedAE)	0.387	0.35	0.46	0.338	0.384	0.387	0.387	0.453	**0.327**	0.347
E1 (MAE)	0.485	0.436	0.572	0.429	0.47	0.485	0.485	0.544	**0.414**	0.417
E2 (RMSE)	0.562	0.588	0.78	0.571	0.61	0.562	0.562	0.622	0.557	**0.556**
E2 (MedAE)	0.351	0.357	0.464	0.344	0.398	0.351	0.351	0.415	**0.334**	0.346
E2 (MAE)	0.433	0.448	0.591	0.437	0.478	0.433	0.433	0.495	**0.422**	0.429
E3 (RMSE)	0.591	0.589	0.785	0.568	0.613	0.591	0.591	0.644	0.554	**0.549**
E3 (MedAE)	0.367	0.362	0.465	0.57	0.375	0.367	0.367	0.433	**0.336**	0.343
E3 (MAE)	0.453	0.45	0.598	0.438	0.466	0.453	0.453	0.512	0.426	**0.417**

$$vk_{t,c} = \begin{cases} D_{JS}(Beta(\alpha_{t,c}, \beta_{t,c}), Beta(\alpha_{t+1,c}, \beta_{t+1,c})), & \frac{\alpha_{t,c}}{\alpha_{t,c}+\beta_{t,c}} < \frac{\alpha_{t+1,c}}{\alpha_{t+1,c}+\beta_{t+1,c}} \\ -D_{JS}(Beta(\alpha_{t,c}, \beta_{t,c}), Beta(\alpha_{t+1,c}, \beta_{t+1,c})), & Sinon \end{cases} \tag{4}$$

$$k_t = \sum_{c=4}^{c=0 \vee k_t \neq 0} \begin{cases} \alpha_{c-1} vk_{t,c-1}; & vk_{t,c} > 0 \\ 0; & Sinon \end{cases} \tag{5}$$

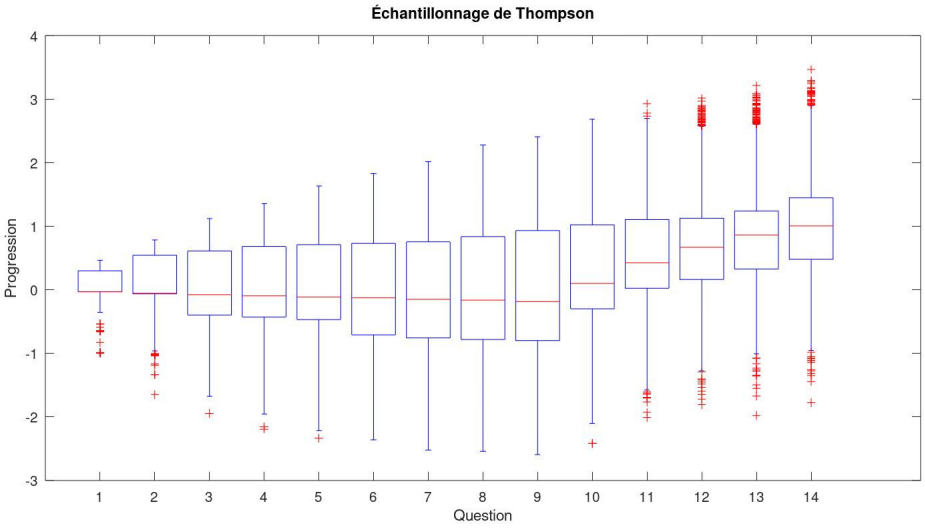


Figure 1. Progression des connaissances avec l'ET selon la divergence de Jensen-Shannon

La figure 1 montre la progression cumulative des connaissances sur quinze questions d'une seule session. Entre la première et la dernière question, tous les apprenants ont statistiquement augmenté leur niveau de connaissance puisque la moyenne a augmenté. La variabilité augmente à partir de la première question jusqu'à la question 9, où le système a acquis plus d'informations sur les apprenants. À ce stade, la variance diminue et la médiane augmente.

Pour le troisième test, la figure 2 montre qu'après 100 exécutions et malgré la stochasticité, l'algorithme est stable : la variance globale dans tous les niveaux de complexité est faible par rapport au nombre total de recommandations.

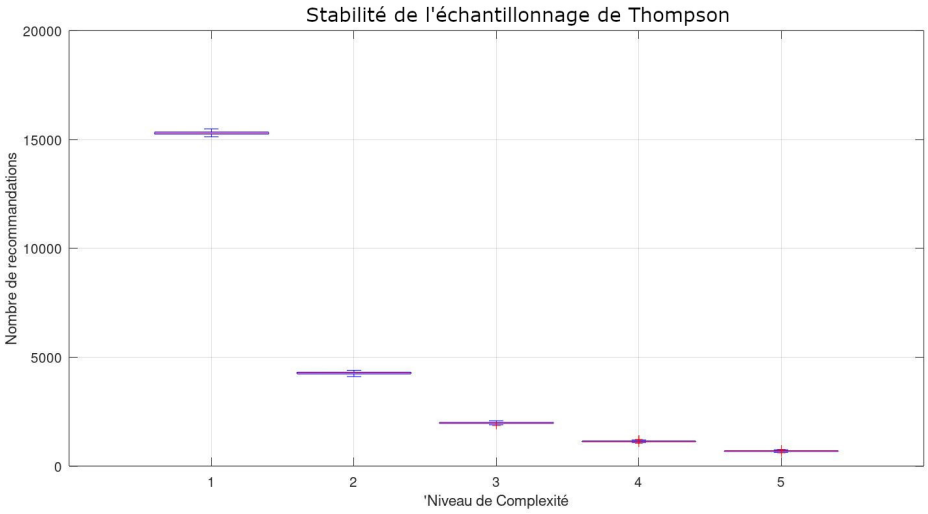


Figure 2. Nombre de recommandations par niveau de complexité

5 Conclusion

Cet article présente l'intégration d'un système de recommandation fondé sur l'échantillonnage de Thompson et un modèle de régression d'ensemble ESCBR-SMA. Le modèle proposé permet de générer des recommandations pour chaque apprenant. Par ailleurs, l'intégration de plusieurs perspectives de solution sur une même problématique, selon le principe du paradoxe de Stein, permet de limiter le risque d'être confronté au paradoxe de Simpson. Pour finir, les deux modèles se complètent mutuellement en améliorant les résultats finaux.

Il est maintenant envisagé d'associer d'autres algorithmes d'intelligence artificielle complémentaires pour bénéficier de plus de données pour l'adaptation.

Références

1. Esmael Ahmed. Student performance prediction using machine learning algorithms. *Applied Computational Intelligence and Soft Computing*, 2024(1) :4067721, 2024.
2. Benjamin Clement, Didier Roy, Pierre-Yves Oudeyer, and Manuel Lopes. Multi-armed bandits for intelligent tutoring systems, 2015.
3. Julia Clemente, Héctor Yago, Javier de Pedro-Carracedo, and Javier Bueno. A proposal for an adaptive recommender system based on competences and ontologies. *Expert Systems with Applications*, 208 :118171, 2022.
4. Jakub Kuzilek, Martin Hlosta, and Zdenek Zdrahal. Open university learning analytics dataset. *Scientific Data*, 4(1) :170171, Nov 2017.
5. Linqing Li and Zhifeng Wang. Knowledge graph-enhanced intelligent tutoring system based on exercise representativeness and informativeness. *International Journal of Intelligent Systems*, 2023(1) :2578286, 2023.
6. Mengchi Liu and Dongmei Yu. Towards intelligent e-learning systems. *Education and Information Technologies*, 28(7) :7845–7876, Jul 2023.
7. Kejie Mao, Qiwen Dong, Ye Wang, and Daocheng Honga. An exploratory approach to intelligent quiz question recommendation. *Procedia Computer Science*, 207 :4065–4074, 2022. Knowledge-Based and Intelligent Information and Engineering Systems : Proceedings of the 26th International Conference KES2022.
8. Daniel Soto-Forero, Simha Ackermann, Marie-Laure Betbeder, and Julien Henriet. Automatic real-time adaptation of training session difficulty using rules and reinforcement learning in the ai-vt its. *International Journal of Modern Education and Computer Science(IJMECS)*, 16 :56–71, 2024.
9. Daniel Soto Forero, Simha Ackermann, Marie-Laure Betbeder, and Julien Henriet. The intelligent tutoring system ai-vt with case-based reasoning and real time recommender models. In *32nd International conference on Case-based reasoning (ICCBR 2024)*, pages 191 – 205, Mérida, Mexico, jul 2024.
10. Daniel Soto Forero, Marie-Laure Betbeder, and Julien Henriet. Ensemble stacking case-based reasoning for regression. In *32nd International conference on Case-based reasoning (ICCBR 2024)*, pages 159 – 174, Mérida, Mexico, jul 2024.
11. Raaa Subha, Naaa Gayathri, Saaa Sasireka, Raaa Sathiyabanu, Baaa Santhiyaa, and Baaa Varshini. Intelligent tutoring systems using long short-term memory networks and bayesian knowledge tracing. In *2024 5th International Conference on Mobile Computing and Sustainable Informatics (ICMCSI)*, volume 0, pages 24–29, 2024.

Conception itérative d'un jeu dans le domaine de la formation médicale

Lyvia, Streich[1,2], Warisara, Suksamran[1,2], Nicolas, Szilas[1], Claire Dupont[2]

[1] TECFA-FPSE, Université de Genève
nicolas.szilas@unige.ch
[2] UDREM, Faculté de Médecine, Université de Genève
claire.dupont@unige.ch

Résumé. Cet article présente la réalisation d'un jeu vidéo pédagogique intrinsèquement intégré destiné à former médecins et internes à l'importance des feedbacks formatifs ainsi qu'à la manière de les dispenser efficacement.
Mots-clés : jeu sérieux, conception itérative, interdisciplinarité, feedback formatif, formation médicale.

Abstract. This article presents the development of an intrinsically integrated educational video game designed to train medical students and residents on the importance of formative feedback and how to deliver it effectively.
Keywords: serious game, iterative design, interdisciplinarity, formative feedback, medical training.

1 La conception de jeux pédagogiques

1.1 Questions d'intégration

Depuis plusieurs décennies, une question centrale anime le domaine des jeux vidéo d'apprentissage : comment intégrer le contenu pédagogique dans un dispositif ludique pour un apprentissage optimal ? La thèse dominante en la matière précise qu'il est essentiel de privilégier une intégration intrinsèque, fusionnant le contenu pédagogique avec les éléments ludiques [1, 3, 9, 14], en particulier avec la mécanique principale du jeu.

Si on peut ainsi définir la notion d'intégration intrinsèque, dans la pratique, concevoir un tel jeu reste un défi. Pourquoi l'intégration intrinsèque est-elle si rare et difficile [7] à mettre en place ? Par définition, la mécanique de jeu doit se fondre avec le contenu d'apprentissage. Donc chaque domaine d'apprentissage nécessite une mécanique de jeu spécifique. Mis à part les rares cas où il existe un jeu dont la mécanique intègre le contenu souhaité, il est souvent nécessaire de créer une nouvelle mécanique de jeu de toute pièce.

Ainsi, dans le domaine des jeux vidéo pédagogiques intrinsèquement intégrés, point de recette universelle, ni jeux cadres [12], ou autre procédure systématique facile à suivre. On retiendra néanmoins quelques principes généraux :

- **Multidimensionnalité** : l'intégration intrinsèque n'est pas à binaire (intégré ou non) mais constitue une notion graduelle et multidimensionnelle [7, 14, 15].
- **Interdisciplinarité** : concevoir un jeu vidéo pédagogique nécessite une étroite collaboration entre experts de disciplines variées (notamment domaine d'apprentissage et game design).
- **Itérativité** : comme tout jeu, un jeu vidéo pédagogique ne peut être réussi du premier coup. Il nécessite de multiples confrontation à des joueurs-euses. Cela résulte de la nature fondamentalement expérientielle et subjective des jeux [5].

L'objectif de cet article est de proposer un cas concret de conception d'un jeu vidéo pédagogique intrinsèquement intégré.

1.2 Contexte de cette étude de cas

Le jeu présenté ci-dessous a été conçu et réalisé en contexte académique, par des étudiant-es en master, dans un contexte d'apprentissage par projet [13]. Bien que réalisé dans un cadre universitaire, la demande initiale et le besoin de formation sont authentiques et émanent d'un commanditaire extérieur. Ce contexte a plusieurs conséquences méthodologiques sur le projet :

- Certaines compétences nécessaires au développement professionnel d'un jeu vidéo (développement technique, graphisme) sont absentes ou assurées de manière "amatrice".
- Jusqu'à la version bêta, la réalisation du jeu repose uniquement sur le temps des étudiants et encadrants et sur l'infrastructure universitaire, sans budget dédié.
- L'implication du commanditaire est modérée, le projet étant gratuit et non critique pour l'organisation qui émet la demande initiale.

Ces contraintes, bien que limitatives, autorisent en même temps une certaine liberté créative et permettent à l'équipe de se focaliser sur l'aspect essentiel du jeu, à savoir la jouabilité et la pédagogie.

2 Domaine d'apprentissage: le feedback médical

Le commanditaire représente l'Unité de Développement et de Recherche en Éducation Médicale (UDREM) de l'Université de Genève, qui a pour but de soutenir le curriculum de formation des médecins au niveau pré-grade, et de former les enseignants qui les accompagnent pour développer leurs compétences. Les médecins en pré-grade bénéficient de l'encadrement de superviseurs, tels que les médecins postgrade et les chefs de clinique, nécessitant un accompagnement pédagogique adapté (voir Fig. 1).

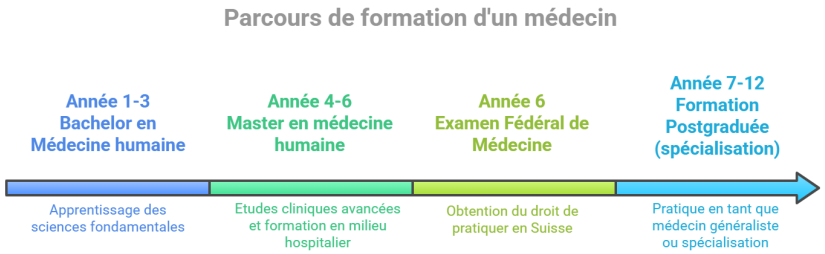


Fig 1. Le parcours de formation des médecins du pré-grade (1 à 6) au post-grade (7 à 12).

Le feedback est un élément déterminant de formation médicale, favorisant le développement des compétences et de la posture médicale.

Il est important , tant pour les apprenants que pour les superviseurs de se former aux techniques de feedback, afin d'être efficaces [16], et de maximiser leurs bénéfices, favorisant ainsi une culture où les points forts et les axes d'amélioration sont abordés sans jugement.

Le socle théorique de ce projet repose sur la méthode de Pendleton [11], qui propose cinq étapes pour un feedback constructif :

1. Auto-évaluation de l'apprenant
2. Discussion et Identification des points positifs observés par l'encadrant et l'apprenant de la supervision directe (observation) ou indirecte (rapport de situation
3. Recherche de solutions et discussion sur les pistes d'amélioration
4. Récapitulatif conjoint du feedback [16]

Cette méthode structure un feedback bilatéral, équilibrant points positifs et améliorations, favorisant réflexion, motivation et apprentissage collaboratif.

Comme le soulignent Henderson et ses collègues [4], un feedback précis et exploitable est essentiel pour permettre aux apprenants de progresser dans leur pratique. Ces auteurs insistent également sur l'importance de former à la fois les superviseurs et les apprenants pour maximiser les bénéfices des retours, favorisant ainsi une culture où les points forts et les axes d'amélioration sont abordés sans jugement.

D'autre part, ce public cible joue un rôle clé dans l'encadrement des stagiaires en formation pré-grade (master en médecine humaine), étant les premiers interlocuteurs hiérarchiques responsables de leur suivi quotidien. L'objectif à long terme est de promouvoir la pratique du feedback à toutes les étapes du parcours de formation des médecins, afin qu'il devienne, tant pour l'encadrant que pour l'apprenant, une habitude naturelle et même un besoin indispensable à leur progression.

3 Le jeu

3.1 Dimension narrative

Le contenu d'apprentissage, axé essentiellement sur les relations interpersonnelles, a orienté le choix vers la conception d'un jeu narratif, où l'histoire joue un rôle

prépondérant. Elle suit un schéma narratif classique de transformation, celle d'une héroïne incompétente en supervision vers une héroïne compétente, ou tout du moins, ayant saisi l'importance de cette dimension.

Par ailleurs, la supervision clinique exige un certain décentrage, soit la capacité à adopter un point de vue différent : l'interne d'aujourd'hui est le stagiaire d'hier et le médecin de demain. Dans ce contexte multi-rôles, il est essentiel de savoir donner et recevoir des feedbacks de manière constructive. L'histoire proposée exploite pleinement ce registre, à travers la thématique du rêve et du changement de rôles qui l'accompagne. Elle intègre de nombreux personnages pour illustrer la complexité des relations entre tous ces rôles professionnels, sans oublier les patients. Si la scène se déroule dans un hôpital, le réalisme n'est pas recherché : le ton utilisé est volontairement décalé, humoristique, allant jusqu'à une parodie du film Matrix, donnant d'ailleurs son nom au jeu, dans l'objectif de rendre l'expérience plus attrayante et marquante contrairement à la majorité des produits du marché perçus comme peu ludiques.

Dans Medtrix [17], les joueurs-euses incarnent Nea, une médecin interne chargée de superviser deux stagiaires. Débordée, elle néglige de leur fournir des feedbacks sur leurs performances, ce qui pousse son chef de clinique à lui reprocher ses manquements. Nea s'endort, dépassée par cette convocation et se réveille dans une simulation de son environnement hospitalier. Très rapidement, elle s'aperçoit qu'elle vit son premier jour en tant que stagiaire au sein de cet hôpital.

Le chef de clinique, Morphy, lui explique qu'elle va devoir accomplir plusieurs missions pour échapper à cette simulation, dénommée "la Medtrice". Il lui présente Trixia : l'interne qui l'accompagne tout au long de son aventure. Elle la guidera aussi à travers son raisonnement en revenant sur les feedbacks exemplaires – ou non – qu'elle recevra lors de ses interactions.

Une fois les missions accomplies, un appel du chef de clinique les conduit à une ultime épreuve : mettre en pratique ses acquis en offrant un feedback constructif à une stagiaire qu'elle supervise.

Lorsque Nea réussit cette dernière épreuve, l'épilogue se déploie, racontant qu'après le retour de son "voyage interdimensionnel", elle mobilise quotidiennement ses compétences en feedback avec les stagiaires qu'elle supervise.



Fig. 2. Première interaction avec le chef de clinique.



Fig. 3. Exemple de mission et aperçu du tableau de bord des joueurs-euses.

3.2 Interactions

Empruntant à l'approche du Challenge-Based Learning [2] la notion de résolution de problèmes, les joueurs-euses débloquent les missions en interagissant avec les PNJ situés dans différents secteurs hospitaliers.

Les dialogues du jeu, inspirés de vignettes cliniques initialement issues de la formation, permettent de faire des choix interactifs, ce qui renforce le sentiment d'agentivité [6] plonge les joueurs-euses dans des situations proches de celles qu'ils rencontreraient dans un contexte professionnel (apprentissage expérientiel de Kolb [8]).

Les dialogues avec Trixia et l'épreuve finale du jeu permettent d'intégrer une phase d'institutionnalisation [10] de la méthode Pendleton en reconsolidant les connaissances acquises lors des missions précédentes.



Fig. 4. Exemple de dialogue avec Trixia, l'interne spécialisée en feedback.

3.3 Développement itératif

Medtrix a été développé avec le logiciel RPG Maker MV [17], selon une approche itérative (prototypage papier puis 4 itérations). Le format et la narration adoptés ont été bien accueillis par le public cible, confirmant la pertinence des choix pédagogiques et scénaristiques. Les retours recueillis lors des itérations ont permis de détecter les incohérences médicales, de fluidifier la navigation et d'améliorer les dialogues.

4. Conclusion

Face à la richesse et la complexité de l'apprentissage du feedback formatif en médecine, nous avons proposé un jeu vidéo pédagogique ambitionnant d'insérer au mieux ce contenu selon le principe de l'intégration intrinsèque. Un tel résultat s'obtient après de multiples itérations et diverses évaluations : évaluations expertes, questionnaires et tests en *play aloud* par des utilisateurs tout venant, stagiaires ou internes en médecine. Cependant, bien que le jeu offre une approche innovante, le format à dialogues et choix limités ne reflète pas pleinement la complexité des interactions réelles en milieu clinique, ce qui pourrait restreindre l'application des compétences acquises dans des situations réelles.

Aujourd'hui, le projet se poursuit, avec des évaluations sur des effectifs élargis et plus ciblées sur le public cible, dans une optique d'amélioration continue avant son déploiement dans une formation existante.

Initialement conçu pour sensibiliser les internes au feedback, Medtrix ouvre aujourd'hui des perspectives d'intégration dans la formation continue. En tant qu'outil pédagogique à part entière, il pourrait contribuer durablement au développement des compétences transversales en santé, au-delà de la seule phase d'initiation.

Références

1. Fabricatore, C.: Learning and Videogames: an Unexploited Synergy. In: Workshop: In Search of the Meaning of Learning. , Long Beach, CA, USA (2000).
2. Gallagher, S.E., Savage, T.: Challenge-based learning in higher education: an exploratory literature review. Teaching in Higher Education. 28, 6, 1135–1157 (2023). https://doi.org/10.1080/13562517.2020.1863354.
3. Habgood, M.P.J. et al.: Endogenous fantasy and learning in digital games. Simulation & Gaming. 36, 4, 483–498 (2005). https://doi.org/10.1177/1046878105282276.
4. Henderson, M. et al.: Conditions that enable effective feedback. Higher Education Research & Development. 38, 7, 1401–1416 (2019). https://doi.org/10.1080/07294360.2019.1657807.
5. Huizinga, J.: Homo ludens. Essai sur la fonction sociale du jeu. Gallimard, Paris (1951).
6. Jézégou, A.: Dictionnaire des concepts de la professionnalisation. In: Dictionnaire des concepts de la professionnalisation. pp. 41–44 De Boeck Supérieur (2022). https://doi.org/10.3917/dbu.jorro.2022.01.0041.
7. Ke, F.: Designing and integrating purposeful learning in game play: a systematic review. Educational Technology Research and Development. 64, 2, 219–244 (2016). https://doi.org/10.1007/s11423-015-9418-1.
8. Kolb, D.: Experiential Learning: Experience As The Source Of Learning And Development. (1984).
9. Malone, T.W.: Toward a Theory of Intrinsically Instruction Motivating. Cognitive Science. 5, 4, 333–369 (1981). https://doi.org/10.1207/s15516709cog0504_2.
10. Margolinas, C.: Dévolution et institutionnalisation: deux aspects antagonistes du rôle du maître. In: Comiti, C. et al. (eds.) Didactique des disciplines scientifiques et formation des enseignants. pp. 342–347 Maison d'Edition de l'Education, Hanoï (1995).
11. Pendleton, D. et al.: The New Consultation: Developing doctor-patient communication. (2003). https://doi.org/10.1093/med/9780192632883.001.0001.
12. Sauvé, L., Samson, D.: Jeux génériques: multiplicateurs de contenu multimédia éducatif canadien sur l'inforoute. SAVIE (2004).
13. Szilas, N.: The mutations of project-based learning in the context of information services. In: Le Dinh, T. and Drăgoicea, M. (eds.) ITM Web of Conferences 41(03002) (2022). https://doi.org/10.1051/itmconf/20224103002.
14. Szilas, N., Acosta, M.: A Theoretical Background for Educational Video Games : Games , Signs , Knowledge. In: Felicia, P. (ed.) Handbook of Research on Improving Learning and Motivation through Educational Games: Multidisciplinary Approaches. IGI Global (2010).
15. Szilas, N., Sutter Widmer, D.: Mieux comprendre la notion d'intégration entre apprentissage et jeu. In: Georges, S. and Sanchez, E. (eds.) Actes de l'atelier jeux sérieux de la 4ème conférence EIAH. pp. 27–39 (2009).
16. Van De Ridder, J.M.M., Wijnen-Meijer, M.: Pendleton's Rules: A Mini Review of a Feedback Method. AJBSR. 19, 1, 19–21 (2023). https://doi.org/10.34297/AJBSR.2023.19.002542.
17. Medtrix, https://tecfa.unige.ch/jeux/medtrix/.
18. RPG Maker MV, https://www.rpgmakerweb.com/products/rpg-maker-mv, (2024).

Index des auteurs

A

Ajouaou Jordan 24
Akretche Idir 344
Alvarez Julian 428, 452
Asswiel Naif 344
Audrey Serna 274

B

Bakhtiari Arman 376
Baussand Lou 240
Benabbou Azzeddine 40
Benlamine Mohamed 400
Bertrand Kuichua Kamdem Ovide 376
Betbeder Marie-Laure 600
Bidaud Camille 366
Blanchard Emmanuel G. 344, 440
Bouamama Madina 354
Boumazguida Karim 140
Broisin Julien 40, 74
Bétrancourt Mireille 240, 590

C

Callahan Christophe 344
Chagrot Marjolaine 516
Chaker Rawad 354
Chassaing-Monjou Adrien 524
Cherbonnier Anthony 516
Cisel Matthieu 58, 366
Colin de Verdière Rémi 388
Colin José 74
Conrad Chloé 92
Cornejo Lupa José Manuel 126

D

D'Ham Cédric 388
De Lièvre Bruno 140, 192
Delestre Nicolas 544
Dequen Gilles 210
Deriaz Marielle 240
Desmarais Michel 376, 576
Dessainjean Fanny 292
Devanne Maxime 416
Di Pietro Melody 192
Diouf Maty 590
Djiezion Guelaybe Enosch 110
Dosso Cheyenne 400
Dufrasne Madeleine 536
Dupont Claire 608

E

El Ayady Anass 416
El Boudouri Yassine 428
El Mawas Nour 256, 416, 524
Essalah Wissal 544
Etcheverry Patrick 126
Ezzaouia Mohamed 560

F

Fleury Anthony 524, 536
Florent Diet 274
Fontaine Benjamin 126
Forest Thierry 292
Forestier Germain 416
Fortin Marie-Noëlle 576

G

Geoffre Trystan 486
George Sébastien 440
Glück Florent 240
Guin Nathalie 92
Guérin Jérôme 516

H

Heintz Christophe 400
Henriet Julien 600
Heutte Jean 516
Hoarau Sébastien 74
Hoogstoel Frédéric 428
Housni Sabrin 140, 192

J

Jamet Eric 226
Jaouadi Mariem 158
Joiron Céline 210
Jolivet Sébastien 172

K

Kessemtini Amine 344
Khan Imran 440
Kirouchenassamy Badmavasan 172
Kumps Audrey 140, 192

L

Lamgarraj Mohamed 210
Lancieri Luigi 428
Le Du Jérémy 452
Le Magadou Maelys 240
Lebis Alexis 256, 464, 474, 524, 536
Leconte Laura 226
Lefevre Marie 92, 308, 582
Legrand 240
Lesbats Jennifer 126
Lescure Emmanuel 324
Luengo Vanda 172

M

Malandain Nicolas 544
Mandran Nadine 158, 256, 464, 474
Marfisi-Schottman Iza 292, 560
Marne Bertrand 292
Marquesuzaà Christophe 126
Marshall-Breton Christopher 554
Martinet Benoît 256
Molinari Gaëlle 590
Morisseau Tiffany 400

N

Nebut Mirabelle 554
Nellis Ezra 292
Nodenot Thierry 126
Nuninger Walter 428

O

Olivier Damien 544

P

Parant Aymeric 210
Parmentier Hélène 570
Perez-Fornos Angélica 240
Peter Yvan 428, 554
Pinos Ullauri Luis Alberto 40
Pissens Camelia 140
Plessis-Ouzariah Inès 486
Poellhuber Bruno 576
Pons Catherine 40
Prior Estelle 158
Pupin Maude 554

R

Raclet Baptiste 324
Rebaudo Sébastien 388
Romero Margarida 498
Ronzon Floriane 576

S

Sadallah Madjid 308, 582
Sanchez Eric 158, 570, 590
Saurel Rémi 324
Schmitt Daniel 452
Schmitt Lucas 344
Seddini Mohammed Oussama 560
Silvestre Franck 324
Simonian Stéphane 354
Soto Forero Daniel 600
Streich Lyvia 608
Strutynska Oksana 498
Suksamran Warisara 608
Szilas Nicolas 608
Sénécail Alain 590

T

Temperman Gaetan 140
Toribio Marta 126
Tremblay Chantal 504

V

Valentin Midez 274
Valentini Gregory 240
Vayre Jean Sébastien 400
Venant Rémi 92
Vermeulen Mathieu 256, 474, 524
Victor Oliva Torres 274

Y

Yessad Amel 172

Élise Lavoué 274

Partenaires



(a) Association des Technologies de l'Information pour l'Education et la Formation



(b) Région Hauts-de-France



(c) Université de Lille



(d) IMT Nord Europe



(e) Polytech Lille



(f) Centre de Recherche en Informatique, Signal et Automatique de Lille (CRIStAL)



(g) Centre Interuniversitaire de Recherche en Éducation de Lille (CIREL)

EIAH

2025 Lille

Environnements Informatiques pour l'Apprentissage Humain

Actes

Actes de la 12ème conférence sur les Environnements Informatiques pour l'Apprentissage Humain

Édités par Élise Lavoué, Margarida Romero et Yvan Peter

Du 10 au 13 juin 2025
Université de Lille, Lille
France

ISBN 978-2-901384-10-6