Actes de la onzième Conférence sur les Environnements Informatiques pour l'Apprentissage Humain (EIAH)

Sébastien Iksal, Chrysta Pélissier, Jean-Marie Gilliot, Patricia Marzin-Janvier

▶ **To cite this version:**

Sébastien Iksal, Chrysta Pélissier, Jean-Marie Gilliot, Patricia Marzin-Janvier. Actes de la onzième Conférence sur les Environnements Informatiques pour l'Apprentissage Humain (EIAH). 2023. ⟨hal-04131278v2⟩

HAL Id: hal-04131278

https://hal.science/hal-04131278v2

Submitted on 3 Jul 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

HAL Authorization

EIAH2023 : 11ème Conférence sur les Environnements Informatiques pour l'Apprentissage Humain

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

12-16 juin 2023 Brest (France)

Actes de la onzième 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 Sébastien Iksal, Chrysta Pélissier,
Jean-Marie Gilliot et Patricia Marzin-Janvier

Du 13 au 16 juin 2023
IMT Atlantique, Brest
France

La onzième Conférence EIAH 2023 a été organisée par l'IMT Atlantique sous l'égide de l'ATIEF
(Association des Technologies de l'Information pour l'Education et la Formation)

Table des matières

1 Comités 7

2 Introduction aux actes de La conférence EIAH'2023 11

3 Le mot du comité d'organisation 16

4 Conférenciers invités 17

5 Table ronde : IA et éducation, à l'heure de ChatGPT 18

6 Les ateliers thématiques participatifs 20

7 Session de communications 1.1 Compréhension et identification de besoins 23

Vers l'identification des phases d'apprentissage procédural en environnement virtuel - *Anaïs Raison, Olivier Augereau, Nathalie Le Bigot, Frédéric Devillers, Sébastien Levieux et Franck Ganier* 24

Conception d'un système de révisions : une étude qualitative exploratoire pour identifier les besoins des enseignants et des élèves - *Denis Ollivier, Franck Silvestre, Jean-Baptiste Raclet, Emmanuel Lescure et Julien Broisin* 36

Etudier l'expérience des visiteurs pour identifier leurs processus d'apprentissage au musée - Vers une conception d'EIAH adaptés à la visite non guidée - *Marine Lagasse, Daniel Schmitt et Bruno De Lièvre* 42

Vers la compréhension des difficultés de lecture en L2 à travers des paramètres acoustiques et de mouvement des yeux - *Sofiya Kobylyanskaya, Ioana Vasilescu, Laurence Devillers et Olivier Augereau* 48

8 Session de communications 1.2 Conception et outils pour la conception 55

Outiller la conception participative et générative de tableaux de bord d'apprentissage - *Madjid Sadallah et Jean-Marie Gilliot* 56

Traçabilité by design : conception d'un système interactif pour améliorer la génération automatique de traces Git pendant une activité d'apprentissage - *Mika Pons, Jean-Michel Bruel, Jean-Baptiste Raclet et Franck Silvestre* 68

co.LAB : une plateforme d'accompagnement pour la conception et l'évaluation collaboratives de jeux destinés à un usage éducatif - *Eric Sanchez, Estelle Prior, Nadine Mandran, Sandra Monnier, Audrey Huguenin, Maxence Laurent, Mariem Jaouadi et Dominique Jaccard* 80

9 Session de communications 2.1 Conception et outils pour la conception - 2ième partie 87

Un framework de conception pour des générateurs d'activités de jeu variées et adaptées - *Bérénice Lemoine, Pierre Laforcade et Sébastien George* 88
MIXAP : Un outil auteur d'activités éducatives en réalité augmentée - *Mohamed Ez-Zaouia, Iza Marfisi-Schottman et Cendrine Mercier* 100
Utilisabilité d'un outil auteur immersif : une étude expérimentale pour la scénarisation de vidéos panoramiques interactives - *Daniel Xuan Hien Mai, Guillaume Loup et Jean-Yves Didier* 112

10 Session de communications 2.2 Pédagogie et ludicisation 120
Le Chaudron Magique : un jeu en Réalité Mixte pour l'apprentissage des fractions en autonomie - *Sofiane Touel, Iza Marfisi-Schottman, Sébastien George, Lucas Hattab et Nicolas Pelay* 121
Détermination de profils relatifs à la mobilisation de schème lors de la résolution de puzzles de programmation - *Marielle Leonard, Maxime Bouton et Yvan Peter* 133
La pédagogie par projet agile abat ses cartes : une métaphore ludique au service de la pratique - *Alexis Lebis, Patrick Delaporte, Romain Deleau, Rémy Pinot et Mathieu Vermeulen* 145

11 Session de communications 3 Modélisation, exploitation et description de ressources pour la conception 153
Modélisation de la maturité numérique des enseignants - État de l'art et conception d'un modèle unifié : MUME - *Christine Michel et Laëtitia Pierrot* 154
Construction et exploitation d'un référentiel de types de tâches d'apprentissage de la programmation - *Sébastien Jolivet, Eva Dechaux, Anne-Claire Gobard et Patrick Wang* 166
Comment décrire les compétences de la pensée informatique dans un jeu sérieux ? - *Mathieu Muratet* 178

12 Session de communications 4.1 Recherche en EIAH : enjeux et méthodologies 192
Environnements informatiques pour l'apprentissage du vocabulaire : une revue systématique de la littérature - *Enzo Simonnet, Mathieu Loiseau et Elise Lavoué* 193
Learning Analytics et recherche scientifique au prisme du droit des données personnelles : contribution d'un projet de recherche en Éducation aux Médias et à l'Information - *Tanja Petelin et Hassina El Kechaï* 205
Favoriser et analyser une auto-récolte des usages du numérique : un enjeu scientifique et méthodologique pour la recherche sur les EIAH - *Marion Paggetti, Viviane Folcher, Emeline Ah-Tchine et Anaïs Loizon* 217
L'Identification des Projets de Logiciel Libre Accessibles aux Nouveaux Contributeurs - *Paul Hervot et Benoit Crespin* 223

13 Session de communications 4.2 Retours d'usage et évaluation 231

Comprendre les usages et effets d'un système de recommandations pédagogiques en contexte d'apprentissage non-formel - *Anaëlle Badier, Mathieu Lefort et Marie Lefevre* 232
Proposition d'Indicateurs d'Écriture Collaborative et leur Évaluation - *Fahima Djelil, Christian Hoffmann, Anis Haddouche, Nadine Mandran et Cedric Dham* 244
Usages dans le temps d'un tableau de bord d'apprentissage dans un jeu sérieux - *Katia Oliver-Quelennec, François Bouchet, Thibault Carron et Claire Pinçon* 256
Vers un outil de débriefing des simulations en réalité virtuelle - *Kelly Minotti, Samir Otmane et Guillaume Loup* 262

14 Session de communications 5
Indicateurs et mesures 270
Mesurer le "flow" lors de l'utilisation d'une plateforme de jeux pédagogiques par des élèves du primaire - *Aous Karoui, Lionel Alvarez, Thierry Geoffre et Quentin Brumeaud* 271
De la modélisation des indicateurs pour l'apprentissage à la capitalisation par les utilisateurs - *Albane Gril, Valérie Renault, Madeth May et Sébastien George* . 283
Caractérisation et mesure des discriminations algorithmiques dans la prédiction de la réussite à des cours en ligne - *Mélina Verger, François Bouchet, Sébastien Lallé et Vanda Luengo* 295

15 Session de posters 307
Instrumentation de l'association de registres sémiotiques dans un assistant de preuve - *Emmanuel Beffara, Martin Bodin, Nadine Mandran et Rémi Molinier* 308
Intégration d'explications dans un référentiel de compétences associé à une plateforme d'apprentissage de l'informatique - *Franck Amadieu, Rémi Venant et Julien Broisin* 312
RLG Model et RLG kit : un modèle et un kit de conception de Role Learning Game - *Gaëlle Guigon, Mathieu Vermeulen, Mathieu Muratet et Thibault Carron* 316
Perspectives d'analyse de commentaires métacognitifs pour la conception d'un EIAH - *Christian Hoffmann, Julien Douady et Nadine Mandran* 320
De l'évaluation classique à l'évaluation adaptative - *Adel Ihichr, Omar Oustous, Younès El Bouzekri El Idrissi et Ayoub Ait Lahcen* 324
Environnement virtuel de simulation dans la formation en Odontologie – Présentation d'un programme de recherche et des résultats d'une étude préliminaire - *Valériane Loison, Fabrice Pirolli, Serena Lopez et Raphaëlle Crétin-Pirolli* . 328
Vers une analyse des pratiques informationnelles des acteurs impliqués dans le domaine des serious games - *Mamoudou Ndiaye, Fabrice Pirolli et Raphaëlle Crétin-Pirolli* 332
Vers une analyse temporelle des processus rédactionnels des élèves - *Léo Nebel, François Bouchet et Denis Alamargot* 336
Analyse du comportement de l'apprenant dans l'apprentissage en ligne et adaptatif - *Omar Oustous, Adel Ihichr, Younès El Bouzekri El Idrissi et Ayoub Ait Lahcen* 340

Système de recommandation de cours basé sur les soft skills : Une approche utilisant les algorithmes génétiques - *Luis Alberto Pinos Ullauri, Alexis Lebis, Abir Karami, Mathieu Vermeulen, Anthony Fleury and Wim Van Den Noortgate* . 344
Impact de la Réalité Virtuelle en formation à distance sur les conditions d'apprentissage - *Laetitia Pluton et Erick Stattner* 348
Analyse d'une stratégie d'autorégulation : La recherche d'aide dans l'apprentissage de la programmation sur une plateforme d'apprentissage en ligne - *Marine Roche, Fahima Djelil, Jérôme Eneau, Jean-Marie Gilliot, Geneviève Lameul et Hugues Pentecouteau* 352
Un dispositif de suivi des apprentissages pour soutenir l'orchestration de la classe - *Madjid Sadallah, Maria Teresa Segarra, Jean-Marie Gilliot et Issam Rebai* 356
De quelques facteurs influençant le sentiment de compétences numériques des enseignant-es en début de carrière - *Denise Sutter Widmer, Cécile Vassaux, Hana Bida, Mireille Bétrancourt et Stéphanie Boéchat-Heer* 360

Index des auteurs 365

Comités

Comité de programme

Président·e·s :

Sébastien Iksal, LIUM, Le Mans Université, France
Chrysta Pélissier, LHUMAIN, Université Paul Valéry, UM3, France

Membres :

Marie-Helene Abel, HEUDIASYC, Université de Technologie de Compiègne, France
Franck Amadieu, University of Toulouse, CLLE UMR5263, France
Francine Athias, ELLIADD-FR EDUC, France
Monique Baron, LIP6, Sorbonne Université, France
Georges-Louis Baron, EDA, Université Paris Cité, France
Vincent Barré, LIUM, Le Mans Université, France
François-Xavier Bernard, EDA, Université Paris Cité, France
Vincent Berry, LIRMM, Université de Montpellier, France
Mireille Bétrancourt, Université de Genève, Suisse
Catherine Bonnat, Université de Genève, TECFA, Suisse
François Bouchet, Sorbonne Université, LIP6, France
Denis Bouhineau, Université Grenoble Alpes, France
Laetitia Boulc'h, Université Paris Cité, France
Jean-François Bourdet, Le Mans Université, France
Nabila Bousbia, LMCS, École nationale Supérieure d'Informatique, Algérie
Julien Broisin, Université Toulouse 3 Paul Sabatier, IRIT, France
Armelle Brun, LORIA, Université de Lorraine, France
Pierre-André Caron, laboratoire CIREL, Université de Lille1, France
Thibault Carron, LIP6, Université Paris 6, France
Ronan Champagnat, L3i, Université de La Rochelle, France
Bernadette Charlier, University of Fribourg, Suisse
Chantal Charnet, Université Paul-Valéry Montpellier, France
Raphaëlle Crétin-Pirolli, Le Mans Université, France
Jean-François Céci, Université de Pau et des Pays de l'Adour, Laboratoire Techné, Poitiers, France
Colin De La Higuera, LS2N, Université de Nantes, France
Bruno De Lièvre, Université de Mons, Belgique
Nicolas Delestre, LITIS, Normandie Université, INSA Rouen Normandie, France
Christian Depover, Université de Mons, Belgique
Michel Desmarais, Ecole Polytechnique de Montréal, Canada
Cyrille Desmoulins, Université Grenoble Alpes, France
Christophe Després, LIUM, Le Mans Université, France

Philippe Dessus, Université Grenoble Alpes, LaRAC, France
Fahima Djelil, IMT Atlantique, LabSTICC, France
Hassina El Kechai, Laboratoire TECHNE, Université de poitiers, France
Nour El Mawas, CIREL, Université de Lille, France
Cédric Fluckiger, Université de Lille 3, 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, University of Grenoble Alpes, France
Monique Grandbastien, LORIA, Université de Lorraine, France
Dominique Groux-Leclet, Université de Picardie, France
Olivier Grugier, INSPE de Paris, France
Nathalie Guin, LIRIS - Université de Lyon, France
Ludovic Hamon, LIUM / Le Mans Université, France
Nathalie Huet, CLLE-LTC, Université Toulouse 2, France
Sébastien Iksal, LIUM, Le Mans Université, France
Stéphanie Jean-Daubias, Université de Lyon, LIRIS, France
Céline Joiron, MIS Laboratory - University of Picardie, France
Sébastien Jolivet, IUFE & TECFA, Université de Genève, France
Jean-Marc Labat, Université Paris 6, France
Yacine Lafifi, Université 8 May 1945 de Guelma, Algérie
Pierre Laforcade, LIUM, Le Mans Université, France
Sébastien Lallé, Sorbonne Université, France
Elise Lavoué, iae lyon, Université Jean Moulin Lyon 3, LIRIS, France
Alexis Lebis, IMT Nord Europe, France
Gaelle Lefer Sauvage Mayotte CUFR de Mayotte, France
Marie Lefevre, LIRIS, Université Lyon 1, France
Dominique Lenne, Heudiasyc, Université de Technologie de Compiègne, France
Pascal Leroux, CREN, Le Mans Université, France
Paul Libbrecht, IUBH Fernstudium, Allemagne
Domitile Lourdeaux, CNRS, Allemagne
Vanda Luengo, Sorbonne Université, LIP6, France
Stéphanie Mailles Viard Metz, Aix Marseille Université, Laboratoire ADEF & Inspé, France
Nadine Mandran, Université Grenoble Alpes, Laboratoire d'Informatique de Grenoble (LIG), 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
Anita Messaoui, LIRDEF, Université de Montpellier, France
Christine Michel, Techne, France
Najoua Mohib, LISEC, Université de Strasbourg, France
Gaëlle Molinari, Formation Universitaire à Distance Suisse (Unidistance) & TECFA (Université de Genève), Suisse
Mathieu Muratet, LIP6, France

Roger Nkambou, Université du Québec à Montréal, Canada
Thierry Nodenot, LIUPPA (EA3000), IUT de Bayonne Pays Basque, France
Sandra Nogry, Laboratoire Paragraphe, Université Cergy-Pontoise, France
Lahcen Oubahssi, LIUM, Le Mans Université, France
Chrysta Pélissier, LHUMAIN, Université de Montpellier 3, France
Claire Peltier, Université Laval, Canada
Yvan Peter, Université de Lille, France
Claudine Piau-Toffolon, LIUM, Le Mans Université, France
Laëtitia Pierrot, Université de Franche-Comté, ELLIADD/NEXT, France
Fabrice Pirolli, Le Mans Université, France
Franck Poirier, Université Bretagne-Sud, France
Françoise Poyet, INSPE de Lyon, France
Issam Rebai, Lab-STICC, IMT Atlantique, France
Christophe Reffay, Université de Franche-Comté, France
Robert Reuter, AES, University of Luxembourg, Luxembourg
Stéphanie Reyssier, Université Lumière Lyon 2, France
Madjid Sadallah, Lab-STICC, IMT Atlantique, France
Eric Sanchez, University of Geneva, Suisse
Yann Secq, Université de Lille, France
Karim Sehaba, LIRIS CNRS, France
Franck Silvestre, Université Toulouse 1 Capitole, IRIT, France
Grégory Smits, Lab-STICC, IMT Atlantique, France
Denise Sutter Widmer, Université de Genève et HEP-Vaud, Suisse
Nicolas Szilas, TECFA-FPSE, University of Geneva, 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, IMT Nord Europe, Institut Mines-Télécom, France
Amel Yessad, LIP6, Sorbonne Université, France

Président·e·s des ateliers :

Madjid Sadallah, Lab-STICC, IMT Atlantique, Brest, France
Denise Sutter Widmer, TECFA-FPSE, Université de Genève, Suisse

Comité d'organisation

Président·e·s :

Jean-Marie Gilliot, Lab-STICC, IMT Atlantique, Brest, France
Patricia Marzin-Janvier, CREAD, Université Bretagne Occidentale, Brest, France

Membres :

Olivier Augereau, Lab-STICC, L'École nationale d'ingénieurs de Brest, France
Matthieu Branthôme, CREAD, Université Bretagne Occidentale, France

Rebecca Clayton-Bernard, CREAD, IMT Atlantique, France
Fahima Djelil, Lab-STICC, IMT Atlantique, France
Charlotte Hoareau, Lab-STICC, Université Bretagne Occidentale, France
Sébastien Kubicki, Lab-STICC, L'École nationale d'ingénieurs de Brest, France
Armelle Lannuzel, IMT Atlantique Bretagne Pays de la Loire, France
Hugues Pentecouteau, CREAD, Université Rennes 2, France
Anais Raison, Lab-STICC, L'École nationale d'ingénieurs de Brest, France
Issam Rebai, Lab-STICC, IMT Atlantique, France
Marine Roche, CREAD, Université Rennes 2, France
Madjid Sadallah, Lab-STICC, IMT Atlantique, France
Grégory Smits, Lab-STICC, IMT Atlantique, France
Maria Teresa Segarra Montesinos, Lab-STICC, IMT Atlantique, France

EIAH2023 : 11ème Conférence sur les Environnements Informatiques pour l'Apprentissage Humain
La conférence pluridisciplinaire francophone sur la conception et l'analyse des environnements numériques pour l'éducation et la formation
12-16 juin 2023 Brest (France)

Introduction aux actes de La conférence EIAH'2023

Sébastien Iksal et Chrysta Pélissier
Président·e·s du comité de programme
Le Mans Université, LIUM, F-72085 Le Mans cedex 9, France
sebastien.iksal@univ-lemans.fr
Université Paul Valéry, LHUMAIN, F-34000 Montpellier, France
chrysta.pelissier@umontpellier.fr

La conférence Environnements Informatiques pour l'Apprentissage Humain (EIAH) est une conférence internationale, pluridisciplinaire et francophone sur la conception et l'analyse des usages des environnements numériques pour l'éducation et la formation, placée sous l'égide de l'ATIEF (Association des Technologies de l'Information pour l'Éducation et la Formation).

La 11ème édition de la conférence EIAH (EIAH'2023) se déroule du 13 au 16 juin 2023 à Brest. Elle succède à la conférence EIAH'2021 qui s'est tenue en juin 2021 en ligne et EIAH'2019 organisée à Paris en juin 2019.

1. L'adaptation comme thème

La thématique scientifique cette année concerne la question de l'adaptation dans le domaine des EIAH, en particulier décrire ou discuter les adaptations technologiques et d'usage(s), leurs caractéristiques et leurs effets sur la construction des connaissances et l'expérience d'apprentissage, mais aussi plus globalement les adaptations qu'elles induisent en fonction des contextes de conception ou d'usages.

Dans le champ de l'éducation, nous pouvons d'abord positionner l'adaptation lors de la conception, à partir de la représentation/modèle de l'élève et/ou de la caractérisation de son développement personnel, de l'école primaire à l'université.

Nous pouvons ensuite l'envisager comme prenant place dans les procédures de développement des EIAH. L'adaptation participe dynamiquement aux évolutions techniques, mais aussi méthodologiques prenant en compte la connaissance scientifique.

Ensuite, l'adaptation prend place dans la mise en œuvre de dispositifs pédagogiques au sein desquels elle impacte les acteurs, mais aussi l'organisation, les choix techniques ou encore pédagogiques.

Enfin, après un usage des utilisateurs (ou plutôt futurs utilisateurs), l'EIAH mais aussi le dispositif dans lequel cet EIAH trouve sa place, peut prendre en compte et donc s'adapter à d'autres connaissances, être intégré dans une nouvelle démarche de conception, de développement ou de mise en œuvre.

Ces adaptations, dans leurs formes, leurs fonctions et leurs niveaux, répondent à un besoin social (humain) et/ou institutionnel (ex : textes officiels), individuel ou collectif, et prennent place dans un dispositif global, se positionnant de manière très spécifique dans l'environnement.

Par l'appel à contribution que nous avons proposé, nous avons souhaité mettre l'accent sur les challenges actuels liés aux nouvelles formes d'apprentissage :

— Comment caractériser et décrire ces nouvelles formes d'adaptation en considérant les aspects techniques, contextuels et humains ?
— Quelles évolutions induisent ces formes d'adaptation en termes de méthodes de conception, de développement, d'évaluation et de conduite de la recherche en EIAH ?
— Quels sont les effets de l'adaptation sur la construction des connaissances et l'expérience d'enseignement/d'apprentissage, et comment les mesurer ?

Pour tenter d'apporter des éléments de réponse à la formalisation d'un paradigme méthodologique et théorique autour de ces questions, les contributions abordent des travaux théoriques ou empiriques visant :

— à favoriser et/ou comprendre une situation d'apprentissage ou d'enseignement, quelles qu'en soient les modalités (synchrone, asynchrone, comodal ; en présence, à distance, hybride ; en formation initiale ou tout au long de la vie ; individuelle ou collective ; formelle/non formelle, etc.),

Et

— à mettre en œuvre un artefact numérique pour soutenir l'apprentissage ou l'enseignement.

2. Une adaptation dans la forme de la conférence

Au-delà des axes de recherche tournés vers l'adaptation, nous avons souhaité adapter notre forme de communication qui caractérise cette conférence EIAH : le mode d'échange entre les participants.

Partant du contexte de l'IMT Atlantique Bretagne-Pays de la Loire, Campus de Brest, tourné vers l'océan mais aussi souhaitant nous faire partager sa culture humaine locale, nous avons souhaité mettre en place une procédure de partage des interactions sociales qui ont lieu habituellement en petits comités, lors des sessions de posters, ou après les différentes sessions de communications ou encore suite aux sessions de conférences invitées.

Ces interactions pourront prendre forme par un partage d'impression de la notion d'adaptation à partir de la mise en place d'un nuage de mots sous la forme d'un Wooclap collectif, laissant apparaître les notions liées à des interprétations, des impressions de chacun des participants, l'objectif

étant d'offrir un moyen direct et convivial d'échanger. Nous avons hâte d'expérimenter ce dispositif et de pouvoir analyser le contenu des différents nuages de mots constitués.

3. Une adaptation des participants aux thématiques de la conférence

En termes d'activité de recherche, 53 articles ont été soumis : 38 articles longs, 6 articles courts, 7 posters et 2 démos.

L'évaluation des propositions s'est faite en quatre étapes :

1. Chaque article soumis a été évalué à la fois par deux relecteurs du domaine de la proposition, et un autre hors domaine ;
2. Ensuite, les auteurs ont pu prendre connaissance des retours de ces deux évaluations et éventuellement apporter un commentaire de réponse ;
3. Puis, un méta-relecteur a animé une discussion avec les deux relecteurs (impliqués dans l'étape 1) afin de proposer une décision d'acceptation ou de rejet ;
4. Enfin, les deux présidents du programme, à la lecture des commentaires laissés par les trois évaluateurs, ont pris une décision définitive.

Ce travail a ainsi impliqué 72 relecteurs et 23 méta-relecteurs.

TABLE 1 – Nombre et type d'articles publiés par rapport aux soumissions

Types de publication	Soumis	Acceptés	Acceptés en court	Acceptés en poster	Rejetés	Retirés par l'auteur	Publiés
Long	38	17	8	8	5	3	17
Court	6	3	/	3	/	1	3
Poster	7	6	/	/	1	/	6
Démo	2	0	/	/	2	/	0
Total	53	26	8	11	8	4	26

Au final, 26 articles ont été publiés : 17 en articles longs, 10 en articles courts et 14 en posters (voir Tableau 1). Le taux de sélection pour les articles longs est de 44%.

TABLE 2 – Répartition des articles publiés dans les domaines EIAH

	Soumissions	Acceptés
SHS	16	13 (81,25%)
Informatique	37	32 (86,48%)

TABLE 3 – Nombre de soumissions et nombre d'acceptation par thématique

Thématiques	Nb soumissions	Nb acceptés	Nb refusés
Evaluation et analyse des effets	16	15	1
Analyse d'usages et de pratiques	13	12	1
Collecte, traitement et analyse des traces d'apprentissage	11	9	2
Jeux éducatifs, jeux sérieux et gamification/ludicisation	10	8	2
Observation et analyse de situations d'apprentissage/de formation	9	9	0
Méthodologie de conception des EIAH	9	8	1
Modalités d'intégration des EIAH dans l'enseignement et la formation	8	8	0
Interfaces et interactions avancées pour l'éducation et la formation	8	7	1
Modélisation de l'apprenant et des interactions	7	7	0
Systèmes adaptatifs	7	6	1
EIAH et modalités d'apprentissage innovantes	7	5	2
Personnalisation et l'apprentissage	6	5	1
Réalités augmentée et virtuelle	6	5	1
Ingénierie pédagogique et scénarisation des EIAH	6	4	2
Systèmes de recommandation	5	5	0
Outils-auteurs, création de ressources pédagogiques	5	4	1
Méthodes d'évaluation des EIAH	5	4	1
Usages en contextes sociaux, éducatifs, culturels ou professionnels spécifiques	5	4	1
Applications favorisant l'apprentissage auto-régulé	4	4	0
Dispositifs d'entraînement, de remédiation et de rééducation	4	4	0
Environnements d'apprentissage collaboratif	4	3	1
Apprentissage automatique et fouille de données	4	3	1
Dimensions didactiques des EIAH	4	3	1
Modalités de collaboration	4	2	2
Modèles et représentations des connaissances pour les EIAH	3	3	0
Apprentissage informel	3	3	0
Tuteurs ou agents intelligents	3	2	1
Techniques de visualisation de données d'apprentissage	3	2	1
Ressources éducatives libres	3	2	1
Evaluation de dispositifs d'apprentissage/de formation	2	2	0
Génération d'activités adaptées et/ou adaptatives	2	2	0
Prise en compte des aspects affectifs et émotionnels dans les EIAH	2	2	0
Simulation, mondes virtuels (3D)	2	2	0
Prise en compte des questions éthiques pour le traitement des données	2	2	0
Ergonomie des interfaces	2	2	0
Traitement automatique des langues en EIAH	2	2	0
Apprentissage hybride	2	1	1
Adaptation culturelle des EIAH	2	1	1
Dimension déontologique de la conception	2	1	1
Modélisation du suivi et de la rétroaction	1	1	0
Web sémantique, ontologies pour les EIAH	1	1	0
EIAH pour les élèves à besoins particuliers	1	1	0
Communautés de pratiques	1	1	0
Prise en compte des questions éthiques pour la collecte et l'archivage des données	1	1	0
Normes et standards pour les EIAH	1	1	0
EIAH ubiquitaires et mobiles	1	1	0
Interopérabilité des ressources, outils et architectures	1	1	0
EIAH et respect de la vie privée	1	0	1

Ce tableau laisse apparaître entre STIC et SHS un taux d'acceptation relativement identique.

Ces différentes soumissions ont été organisées autour de 7 grandes thématiques qui correspondent aux sessions de la conférence :

1. Compréhension et l'identification des besoins ;
2. Recherche en EIAH : enjeux et méthodologies ;
3. Conception et outils pour la conception ;
4. Pédagogie et ludicisation ;
5. Modélisation, exploitation et description de ressources pour la conception ;
6. Indicateurs et mesures ;
7. Retour d'usages et évaluation.

Les caractéristiques des EIAH qui sont plus spécifiquement travaillées sont, selon une dynamique qui se développe depuis plusieurs années autour des jeux sérieux, de l'exploitation des traces d'apprentissage ou encore des modalités de conception d'EIAH intégrant la réalité virtuelle, augmentée, mixte mais aussi les référentiels. Ces constats restent assez cohérents avec les éditions 2017, 2019 et 2021.

Deux nouvelles thématiques ont cependant fait leur apparition cette année : une réflexion sur le droit des données personnelles et l'identification des projets de logiciels libres.

4. Remerciements

Nous tenions à remercier le comité d'organisation des établissements : IMT Atlantique, UBO, ENIB, Université de Rennes 2, qui s'est proposé pour nous accompagner dans la mise en place de cette 11° édition de la conférence EIAH'2023 et la prise en charge des conférenciers invités.

Nous remercions également tout particulièrement les membres du comité de programme, pour la qualité de leurs expertises, leur engagement et leur réactivité. Les retours, conseils et avis donnés aux auteurs lors de la phase de rédaction ont été extrêmement constructifs pour affiner les contributions scientifiques et participer à la qualité de cette conférence.

Nous tenons enfin à remercier l'ATIEF qui nous a fait confiance dans l'organisation de la conférence, soutenue dans les différentes étapes qui nous ont permis encore une année d'assurer la qualité scientifique des publications.

Sur le plan scientifique, la conférence EIAH'23 a bénéficié d'une réflexion de fond, menée en concertation avec la revue STICEF, sur les modalités d'évaluation des articles scientifiques. Les consignes et la grille d'évaluation utilisées pour la sélection des travaux ont permis aux relecteurs du comité de programme de produire des évaluations précises et constructives pour les auteurs, mais aussi pour les méta-relecteurs qui ont échangé avec les relecteurs afin d'arriver à un consensus de décision.

Enfin, remercions les deux conférenciers invités pour leur présence à EIAH 2023 ainsi que le jury du Prix Monique Granbastien.

EIAH2023 : 11ème Conférence sur les Environnements Informatiques pour l'Apprentissage Humain
La conférence pluridisciplinaire francophone sur la conception et l'analyse des environnements numériques pour l'éducation et la formation
12-16 juin 2023 Brest (France)

Le mot du comité d'organisation

Jean Marie Gillot et Patricia Marzin-Janvier
Lab-STICC, IMT Atlantique, Brest, France
jm.gilliot@imt-atlantique.fr
CREAD, Université Bretagne Occidentale, Brest
patricia.marzin-janvier@inspe-bretagne.fr

L'idée de l'organisation de la conférence EIAH2023 sur le site de Brest de IMT Atlantique a été lancée lors d'un Comité d'Administration de l'ATIEF en mai 2022. L'engagement des chercheurs locaux a été immédiat et leur enthousiasme ne s'est pas démenti depuis. La dimension du comité d'organisation en témoigne.

Accueillir à Brest, le colloque francophone sur les Environnements Informatiques d'Apprentissage Humain, au croisement entre l'informatique, les sciences de l'éducation et les sciences du langage, est apparu comme une évidence. Les équipes spécialisées en EIAH et en Environnements Mixtes du Lab-STICC, les équipes du CREAD, en didactique des disciplines et sciences de l'éducation travaillent régulièrement ensemble, notamment au travers de l'organisation du colloque "Questions de Pédagogies dans l'Enseignement Supérieur", né justement à Brest en 2001.

L'équipe organisatrice de la 10e Conférence, à Friburg, en juin 2021 du fait de la pandémie nous avait proposé une plage virtuelle pour nous retrouver. Elle est cette année bien présente au pied du lieu d'accueil de la conférence. Pour joindre l'utile et l'agréable, nous profiterons en outre d'une excursion en mer ou à pied en bord de mer. Nous nous retrouvons également pour un temps convivial à Océanopolis, centre de culture scientifique dédié à l'Océan. Le site de l'IMT Atlantique, outre sa position sur le bord de mer, a été retenu de par sa disponibilité et de par la qualité de son infrastructure qui permet le meilleur accueil pour un tel événement.

Nous remercions l'ATIEF, de nous avoir accordé sa confiance, et aux président.e.s du comité de programme pour leurs interactions qui ont facilité notre travail. Les deux laboratoires d'accueil de ce colloque portent quotidiennement cette interdisciplinarité à la croisée de l'informatique et des sciences humaines. Nous remercions les directrices du CREAD, Patricia Marzin-Janvier et Geneviève Lameul, et les directeurs du Lab-STICC, Christian Person et Philippe Coussy, pour leur soutien et leur confiance. Les services de l'IMT Atlantique en tête, de l'Université de Bretagne Occidentale et de l'ENIB également nous ont apporté leur soutien logistique précieux et nécessaire à l'organisation de la conférence. Qu'ils en soient également remerciés.

L'accueil d'un tel colloque est une belle forme de reconnaissance pour les équipes locales, c'est du moins dans cet esprit que nous avons travaillé collectivement. Nous espérons que vous apprécierez votre venue à Brest à la mesure de l'engagement de l'équipe d'organisation. Qu'elles et ils soient ici toutes et tous remerciés pour leur investissement.

EIAH2023 : 11ème Conférence sur les Environnements Informatiques pour l'Apprentissage Humain
La conférence pluridisciplinaire francophone sur la conception et l'analyse des environnements numériques pour l'éducation et la formation
12-16 juin 2023 Brest (France)

Conférenciers invités

Comprendre les tuteurs intelligents et leur apport pour le milieu éducatif

Valéry Psyché, professeure régulière à l'Université TÉLUQ
valery.psyche@teluq.ca

Dans le cadre de ce séminaire, la professeure Psyché présente sa vision du domaine de l'intelligence artificielle en éducation (mieux connu sous AIED pour Artificial Intelligence in EDucation). Elle aborde l'objet principal de la recherche en lien avec ce domaine bien établi depuis une cinquantaine d'années : les systèmes tutoriels intelligents ou communément appelés, tuteurs intelligents. En abordant le sujet "Comprendre les tuteurs intelligents et leur apport pour le milieu éducatif", elle présentera cet objet de recherche de façon théorique et pratique (exemples à l'appui). Puis, elle fera le lien avec le milieu de l'éducation, ce dernier ayant des préoccupations parfois éloignées de celles de la recherche, en répondant à la question de l'apport possible de la recherche en AIED pour l'écosystème éducatif particulièrement avec les tuteurs intelligents. Pour finir, elle abordera les défis et perspectives pour le système éducatif, notamment par rapport à la transformation du rôle de l'enseignant et de la formation due à l'omniprésence de plus en plus forte de l'IA dans le système éducatif.

Soutenir l'autonomie des apprenants dans les EIAH avec des *prompts* et des *feedbacks*

Franck Amadieu, Professeur de psychologie cognitive et ergonomie
au laboratoire CLLE de l'Université de Toulouse
franck.amadieu@univ-tlse2.fr

Les EIAH peuvent exiger une autonomie importante chez les apprenants, en particulier en situation d'apprentissage à distance et en l'absence de tuteurs humains. Ces exigences d'autonomie se traduisent par la nécessité pour l'apprenant de mettre en œuvre les processus mentaux adéquats pour son apprentissage. Ces processus impliquent la sélection, l'organisation et l'intégration en mémoire des informations pertinentes, mais aussi, la réalisation des procédures pertinentes ainsi que la mise en œuvre de processus métacognitifs pour l'autorégulation de son apprentissage. Afin d'aider les apprenants dans cet objectif d'autonomie avec les EIAH, les systèmes de prompts (incitations) et de feedback apparaissent tout particulièrement utiles. Des travaux dans le cadre du projet DIMEDD ont par exemple confirmé certains effets des prompts durant l'activité de visionnage de vidéo dans un EIAH. En ce qui concerne les feedbacks proposés par un EIAH, ils sont de différentes natures et peuvent aider à accompagner efficacement les apprenants dans leur apprentissage. Les recherches doivent être poursuivies sur les effets des prompts et des feedbacks, comme par exemple sur l'explicabilité de ces derniers.

EIAH2023 : 11ème Conférence sur les Environnements Informatiques pour l'Apprentissage Humain
La conférence pluridisciplinaire francophone sur la conception et l'analyse des environnements numériques pour l'éducation et la formation
12-16 juin 2023 Brest (France)

Table ronde : IA et éducation, à l'heure de ChatGPT

Jean-Marie Gilliot et Hugues Pentecouteau
Lab-STICC, IMT Atlantique, Brest
jm.gilliot@imt-atlantique.fr
CREAD, Université de Rennes 2
hugues.pentecouteau@univ-rennes2.fr

Quels enjeux, quels effets ?

L'arrivée de ChatGPT créé par OpenAI en novembre 2022, qui permet de générer toute forme de texte et de converser avec les humains sur n'importe quel sujet, peut fasciner mais peut aussi inquiéter. Pour l'éducation, certains voient des nouveaux outils incontournables en classe, mais d'autres craignent une triche généralisée.

Nous profitons de la conférence EIAH 2023 qui regroupe de nombreux chercheurs spécialistes en environnements numériques pour l'éducation et la formation pour proposer une table ronde permettant d'aborder les questions soulevées par l'usage de l'IA en éducation : Quels nouveaux rapports au savoir ? La pédagogie va-t-elle évoluer ? Y-a-t-il de nouvelles compétences à acquérir ? Quels seront les outils qui auront leur place dans la formation ?

Cette table ronde laissera une large place aux questions du public pour permettre un débat ouvert.

Participants

La table ronde sera animée par Jean-Marie Gilliot, maître de conférences en informatique au Lab-STICC/ IMT Atlantique et Hugues Pentecouteau, professeur Professeur des universités en sciences de l'éducation et de la formation au CREAD/ université de Rennes 2, et fera intervenir :

— Valéry Psyché, Professeure à l'université TELUQ, Montréal
— Franck Amadieu, Professeur de psychologie cognitive et ergonomie, CLLE, Université de Toulouse
— Marie Lefèvre, Maître de conférences en informatique, LIRIS, Université Claude Bernard - Lyon 1
— Rémi Venant, Maître de conférences en informatique, LIUM, Le Mans Université

Informations pratiques

Cette table ronde pour réfléchir ensemble et avec le public se déroule le jeudi 15 juin à 18h 45 au pôle numérique du Bouguen, à l'UBO à Brest, et est retransmise en ligne (`https://youtube.com/live/CbWb7S8Weag?feature=share`).

Pôle numérique du Bouguen : 6 rue du Bouguen, Brest.

EIAH2023 : 11ème Conférence sur les Environnements Informatiques pour l'Apprentissage Humain
La conférence pluridisciplinaire francophone sur la conception et l'analyse des environnements numériques pour l'éducation et la formation
12-16 juin 2023 Brest (France)

Les ateliers thématiques participatifs

Madjid Sadallah et Denise Sutter Widmer
Lab-STICC, IMT Atlantique, Brest, France
madjid.sadallah@imt-atlantique.fr
TECFA-FPSE, Université de Genève, Suisse
denise.sutterwidmer@unige.ch

Les ateliers thématiques participatifs de la conférence EIAH 2023, qui ont eu lieu le 13 juin, ont été l'occasion de favoriser des échanges enrichissants parmi les participants. Ces ateliers ont joué un rôle essentiel dans le renforcement des partenariats existants et l'établissement de nouvelles collaborations interdisciplinaires. Les participants ont proposé des ateliers selon deux formats : les symposiums, qui étaient sur invitation, et les workshops, qui impliquaient la soumission d'articles. Quelle que soit la formule adoptée, l'objectif primordial était de promouvoir l'interdisciplinarité et de solliciter des propositions provenant de divers domaines disciplinaires et d'institutions variées. Parmi les onze soumissions reçues, sept symposiums (dont un en deux parties) et trois workshops ont été sélectionnés. Un atelier a dû être annulé en raison d'un nombre insuffisant de participants. Les thématiques abordées ont couvert un large éventail de sujets, témoignant de la diversité des approches, des domaines disciplinaires et des parties prenantes impliquées. Les ateliers ont abordé à la fois des thématiques originales et émergentes dans le domaine des EIAH, ainsi que des sujets plus traditionnels revisités selon des perspectives nouvelles.

Quatre grandes thématiques ont émergé cette année lors de la conférence EIAH. La première concerne la conception et l'évaluation des EIAH, abordées à travers le suivi éducatif via des tableaux de bord, l'utilisation des outils et les retours d'expérience dans le contexte des compétences, l'identification des obstacles dans la conception de jeux éducatifs, ainsi que la création et l'évaluation d'une plateforme pour accompagner la recherche orientée par la conception de jeux numériques. La deuxième thématique porte sur l'enseignement du numérique et l'apprentissage dans les EIAH, avec un focus sur l'apprentissage de la pensée informatique et l'impact des facteurs émotionnels et psychologiques sur les apprenants. Une troisième thématique, davantage ancrée dans une démarche citoyenne, met en lumière des enjeux sociétaux et éthiques tels que la sobriété numérique et l'adoption et la pérennisation des logiciels libres. Enfin, la quatrième thématique concerne la recherche en EIAH qui met en avant la diversité et la richesse des approches méthodologiques et théoriques, ainsi que la nécessité de développer un socle de connaissances communes au sein de la communauté EIAH.

1. Conception et évaluation des EIAH

Le symposium *Les tableaux de bord de suivi en éducation* s'est concentré sur les tableaux de bord en éducation, l'élaboration d'indicateurs, l'analyse des données d'apprentissage (Learning Analytics) et la démarche centrée utilisateurs. Il était organisé par Christian Hoffmann (Université Grenoble Alpes) et Fahima Djelil (IMT Atlantique, Brest).

Le symposium *Approche par compétences et EIAH : partage d'outils et retours d'usages*, organisé par Nathalie Guin (Université Lyon 1) et Christine Michel (Université de Poitiers), a mis l'accent sur la modélisation des compétences, le diagnostic des compétences de l'apprenant, l'auto-régulation, la personnalisation et l'analyse des usages.

Le workshop *Obstacles et besoins de la ludopédagogie a examiné les jeux éducatifs* à travers trois axes thématiques : 1) La littératie ludique comme obstacle, 2) L'immersion, l'engagement et l'apprentissage et 3) La ludopédagogie, la recherche et l'éthique. Organisé par Fanny Barnabé (Université de Namur) et Antoine Taly (CNRS - Institut de Biologie Physico-Chimique), il avait pour objectif de recenser les éléments faisant obstacle à la conception de jeux éducatifs ainsi que d'identifier les besoins des acteurs du champ.

Le symposium *Conduire une recherche collaborative orientée par la conception dans le domaine des jeux numériques* était organisé par Estelle Prior (Université de Genève), Mariem Jaouadi (Université de Genève), Nadine Mandran (Université Grenoble-Alpes), Eric Sanchez (Université de Genève) et Dominique Jaccard (HES-SO). Il s'est intéressé à la conduite de la recherche orientée par la conception (ROC), la conception des EIAH, l'évaluation des EIAH et les jeux. Il vise à discuter des obstacles rencontrés, des méthodes et outils nécessaires à la conduite de la ROC lors de la conception de jeux numériques et à développer un cadre méthodologique commun dans la communauté EIAH.

2. Apprentissage du numérique et facteurs humains impactant l'apprentissage

Le workshop *Apprentissage de la pensée informatique de la maternelle à l'Université* organisé par le Groupe de Travail APIMU (Apprentissage de l'Informatique de la Maternelle à l'Université) avec le projet ANR IE CARE (Apprentissage de l'Informatique de la Maternelle à l'Université) était porté par Yvan Peter (Université de Lille), Yannick Parmentier (Université de Lorraine), Yann Secq (Université de Lille), Julien Broisin (Université de Toulouse), Christophe Declercq (INSPÉ de la Réunion) et Cédric Fluckiger (Université de Lille). Il avait pour objectif d'explorer la pensée informatique, la didactique de l'informatique, l'apprentissage de la programmation, l'analyse des données et l'apprentissage collaboratif.

Le workshop *Éducation 4.0 : caractérisation des facteurs humains dans les parcours pédagogiques* s'est intéressé à l'automatisation de l'ensemble des processus pédagogiques et à l'impact des facteurs liés à l'humain et à l'environnement social et technique sur le parcours des apprenants. L'équipe organisatrice de l'atelier était composée de Philippe Aniorté (Indicatic, Panama), Clément Dussarps (Université de Bordeaux) et Laurent Gallon, LIUPPA (Université de Pau et des Pays de l'Adour).

3. Démarche citoyenne, enjeux éthiques, sociétaux et environnementaux

Le symposium *Vers la sobriété numérique dans les EIAH – Partie I et Partie II* était organisé par Nicolas Szilas (Université de Genève), Dominique Py (Université du Mans, Djamileh Aminian (Université de Genève & Société e-teach) et Gaëlle Molinari (Université de Genève & UniDistance). Il abordait la sobriété numérique, l'éco-conception et les moyens de réduire l'empreinte environnementale des EIAH.

Le symposium *Libérer les EIAH grâce aux communs et aux Logiciels Libres*, organisé par Bertrand Marne (LIUM) et Fabrice Pirolli (CREN), avait pour but d'explorer comment le développement, l'utilisation, la diffusion des logiciels libres peuvent promouvoir la construction d'un commun et assurer l'adoption et pérennisation des EIAH.

4. Diversité de la recherche en EIAH

Le symposium *La diversité de la recherche en EIAH par l'exemple* s'est intéressé aux attentes des différents champs disciplinaires de la communauté EIAH au niveau des théories, des contributions et des méthodes de travail afin de constituer un socle de base issu des différentes disciplines. Il était organisé par Franck Silvestre (IRIT, Toulouse) et Nadine Mandran (LIG, Grenoble).

EIAH2023 : 11ème Conférence sur les Environnements Informatiques pour l'Apprentissage Humain
La conférence pluridisciplinaire francophone sur la conception et l'analyse des environnements numériques pour l'éducation et la formation
12-16 juin 2023 Brest (France)

Session de communications 1.1
Compréhension et identification de besoins

Vers l'identification des phases d'apprentissage procédural en environnement virtuel

Anaïs Raison[1], Olivier Augereau[2], Nathalie Le Bigot[1], Frédéric Devillers[2], Sébastien Levieux[1], Franck Ganier[1]

[1] Univ Brest, Lab-STICC, CNRS, UMR 6285, F-29238 Brest Cedex 3, France
anais.raison@univ-brest.fr
nathalie.lebigot@univ-brest.fr
sebastien.levieux@etudiant.univ-brest.fr
franck.ganier@univ-brest.fr
[2] ENIB, Lab-STICC, CNRS, UMR 6285, F-29238 Brest Cedex 3, France
augereau@enib.fr
devillers@enib.fr

Résumé. Les environnements virtuels utilisés pour acquérir des compétences génèrent de grandes quantités de données comportementales et apportent ainsi de nouvelles opportunités de compréhension de l'activité des apprenants. A partir de telles données, une méthode de *clustering* a été utilisée dans la présente étude afin d'identifier des phases dans le processus d'apprentissage procédural. La reconnaissance de ces phases d'un point de vue comportemental permettrait de suivre en temps réel et de façon automatique la progression des individus et ainsi personnaliser les scénarios pédagogiques en environnement virtuel. Soixante-trois participants ont réalisé de façon répétée une procédure d'assemblage du cube de Soma en environnement virtuel. Les analyses montrent une amélioration des performances et une diminution du coût cognitif au fur et à mesure des répétitions de la tâche. Les résultats d'une approche multidimensionnelle de découpage de l'activité en différentes phases sont présentés et discutés.

Mots-clés : Apprentissage procédural, charge de travail mental, données comportementales, environnement virtuel, phases d'apprentissage.

Abstract. Virtual environments used to acquire skills generate large amounts of behavioral data and thus provide new opportunities to understand learners' activity. From such data, a clustering method was used in the present study to identify phases in the procedural learning process. The recognition of these phases from a behavioral point of view would allow to follow in real time and in an automatic way the progress of the individuals and thus to personalize the pedagogical scenarios in a virtual environment. Sixty-three participants repeatedly performed a Soma cube assembly procedure in a virtual environment. Analyses show an improvement in performance and a decrease in cognitive cost as the task is repeated. The results of a multidimensional approach of dividing the activity into different phases are presented and discussed.

Keywords: Procedural learning, mental workload, behavioral data, virtual environment, learning phases.

1 Introduction

La pratique, activité essentielle pour apprendre et progresser dans les domaines industriel, médical ou encore sportif, est caractéristique de l'apprentissage procédural [1, 2]. Ce processus dynamique, parfois assimilé à l'acquisition de compétences ou d'habiletés, s'illustre par une progression du comportement [3] et une évolution des processus cognitifs et des zones cérébrales impliqués au fur et à mesure des répétitions de la tâche [4]. Quel que soit le domaine de compétences, l'évolution des apprenants reste généralement mesurée au travers de données comportementales ou de performance. Les progrès technologiques et le développement de domaines tels que le *learning analytics* [5] et le *e-learning* [6] montrent l'importance de ces données. Le suivi de l'activité des apprenants est en effet facilité grâce aux Environnements Virtuels d'Apprentissage Humain (EVAH) qui offrent la possibilité d'enregistrer automatiquement de nombreuses données comportementales [7] tout en réduisant les coûts et le danger inhérent à certaines formations [3]. Les données quantitatives présentent notamment l'avantage d'être recueillies facilement et en temps réel, c'est pourquoi elles ont déjà été utilisées à des fins de prédiction des performances et de l'apprentissage [8] ou dans le but d'adapter la formation aux apprenants avec par exemple la génération de *feedback* [7].

Afin de pouvoir répondre à la problématique de l'adaptation personnalisée de la situation d'apprentissage en environnement virtuel, la présente étude s'intéresse à l'analyse des traces de l'activité lors d'une situation d'apprentissage procédural par traitement d'instructions [9]. L'enjeu est d'élaborer une méthode basée sur le recueil de données comportementales permettant de repérer dans quelle phase d'apprentissage se situe un individu lorsqu'il utilise un EVAH en vue d'individualiser les scénarios pédagogiques. En effet, les travaux réalisés dans le domaine de l'apprentissage de procédure montrent que ce type d'apprentissage se déroule en plusieurs phases [4, 10–12]. La présente étude repose ainsi sur l'hypothèse que certains indicateurs comportementaux sont susceptibles de refléter les processus cognitifs sous-jacents à l'apprentissage de procédure [3, 10]. Le but est d'évaluer une approche de *clustering* pour tenter d'identifier des phases lors d'un apprentissage procédural par traitement d'instructions en examinant les regroupements de profils comportementaux.

L'article est structuré comme suit. La section 2 présente un état de l'art des théories et des travaux concernant la distinction de trois phases lors d'un apprentissage procédural. La section 3 présente une étude menée pour évaluer l'apprentissage procédural en environnement virtuel et examiner une approche de distinction des phases d'apprentissage. Les résultats de cette étude sont présentés en section 4 et discutés en section 5.

2 État de l'art

L'apprentissage procédural est un processus dynamique qui a fait l'objet de nombreuses modélisations le découpant en trois phases. Le travail de Fitts et Posner sur le développement d'un modèle d'acquisition d'habiletés motrices sert souvent de référence [11]. Ce modèle en trois phases, nommées cognitive, associative et autonome permet d'expliquer la transformation progressive de connaissances déclaratives en connaissances procédurales [11]. Anderson a repris cette distinction en différentes phases dans une théorie nommée ACT-R, pour *Adaptive Control of Thought Rational* [10 ; 13]. Il l'a élargie aux habiletés cognitives et a renommé les phases par : déclarative, transitoire et procédurale. Rasmussen, également influencé par les travaux de Fitts et Posner [11], a proposé en 1983 un modèle, nommé SRK (*Skill-Rule-Knowledge*) qui différencie l'exécution des tâches comme basée soit sur les connaissances, soit sur les règles, soit sur les compétences [14]. Enfin, VanLehn a décrit trois phases d'acquisition des habiletés cognitives dont les caractéristiques se rapprochent de celles du modèle de Fitts et Posner [11] : les phases précoce, intermédiaire et tardive [15]. Ainsi, malgré des dénominations de phases différentes, ces auteurs fournissent une compréhension consensuelle de l'apprentissage puisque tous proposent un processus d'apprentissage en trois phases.

À partir de ce découpage de l'activité des apprenants, plusieurs méthodes d'identification des phases d'apprentissage ont émergé. Par exemple, Taylor et Ivry [16] indiquent que pour des tâches de différentes natures, les courbes d'apprentissage présentent une forme générale en adéquation avec le modèle de Fitts et Posner [11]. Néanmoins, ce lien entre des indicateurs graphiques et des phases d'apprentissage propose une identification *a posteriori* des phases et une délimitation relativement floue de celles-ci. Par ailleurs, certains auteurs abordent l'importance de distinguer la courbe d'apprentissage basée sur des données moyennées de celle basée sur des données individuelles puisqu'elles ne présentent pas la même forme [17]. Ainsi, une courbe d'apprentissage élaborée à partir de données moyennes serait lissée, alors qu'une courbe d'apprentissage élaborée à partir de données individuelles présenterait un profil en dents de scie.

D'autres approches d'identification des phases d'apprentissage ont été proposées dont celles de Ganier et al. [3] et Hubert et al. [12] qui proposent des indicateurs comportementaux et des règles associées pour déterminer les transitions d'une phase à l'autre de manière individuelle. Ils ont ainsi émis l'hypothèse autour d'un indicateur comportemental unique pour décomposer l'activité en trois phases : le nombre de mouvements pour Hubert et al. [12] et le nombre de consultations de consignes pour Ganier et al. [3]. Ces travaux offrent un moyen intéressant de suivre en temps réel l'évolution de chaque apprenant dans les trois phases d'apprentissage. Il faut toutefois rester prudent car un seul indicateur comportemental peut ne pas refléter l'ensemble du processus d'apprentissage.

Plus récemment, l'existence de ces différentes phases d'apprentissage a pu être avérée par des techniques d'imagerie cérébrale (IRM Fonctionnelle et Tomographie par Emission de Positons ; [4 ; 12 ; 18]). En effet, ces travaux ont révélé des changements de niveau d'activation des aires cérébrales au cours des exécutions répétées d'une tâche pouvant refléter le passage d'une phase d'apprentissage à l'autre.

Pour conclure, ces travaux permettent d'identifier les trois phases de l'apprentissage à partir de données comportementales ou de données d'imagerie cérébrale, et ce principalement dans le cas d'une situation d'apprentissage par résolution de problèmes. Toutefois, ces travaux utilisent des méthodologies parfois lourdes et invasives et ne s'appuient pas sur la combinaison de données comportementales pour découper l'apprentissage en diverses phases. Or, ces données offrent la possibilité d'une détection de l'évolution des apprenants en temps réel, sans méthode intrusive, ce qui permettrait de personnaliser le scénario pédagogique selon la phase d'apprentissage atteinte. Le travail qui suit présente donc une approche pour tenter d'identifier des phases lors d'un apprentissage procédural par traitement d'instructions en s'appuyant sur plusieurs données comportementales enregistrées automatiquement par l'EVAH. Plusieurs hypothèses ont été formulées : • H1 : une exécution répétée de la tâche devrait induire de meilleures performances jusqu'à l'atteinte de performances optimales stables • H2 : la charge de travail mental devrait être élevée au début de l'apprentissage puis diminuer au fur et à mesure des exécutions de la tâche pour enfin aboutir à une absence de fluctuations • H3 : des indicateurs comportementaux devraient permettre d'identifier des phases de l'apprentissage procédural.

3 Expérience

Afin d'identifier les phases d'apprentissage de procédure en EVAH à partir d'indicateurs comportementaux recueillies automatiquement par l'environnement virtuel, une expérience a été menée au Centre Européen de Réalité Virtuelle (CERV) situé au Technopôle Brest-Iroise.

3.1 Participants

Soixante-trois individus ont été recrutés pour participer à l'expérience (18 femmes et 45 hommes). Ils étaient âgés de 24 ans en moyenne (σ = 2,58 ; *min* = 19 ; *max* = 32). Tous avaient atteint au minimum le niveau d'étude du baccalauréat et avaient poursuivis dans des spécialisations différentes. Certaines variables ont été mesurées comme l'utilisation antérieure de casques de Réalité Virtuelle (RV) ou encore la pratique des casse-têtes et la connaissance du cube de Soma. Par ailleurs, la cinétose a été évaluée grâce au *Simulator Sickness Questionnaire* (SSQ) [19] administré au début et à la fin de l'expérience et aucun participant n'a déclaré avoir ressenti d'effets secondaires à la suite de l'expérience. Préalablement à la réalisation de l'expérience, les participants ont signé un formulaire de consentement éclairé expliquant notamment qu'ils pouvaient se retirer de l'étude à tout moment, que leur anonymat était préservé et ont été informés des risques secondaires potentiels de la RV.

3.2 Tâche et Environnement Virtuel d'Apprentissage Humain

L'Environnement Virtuel d'Apprentissage Humain (EVAH), développé à partir du moteur *Unity 3D*, représentait un atelier avec des tables et des étagères. Un interrupteur

était présent au centre de l'Environnement Virtuel (EV) : il permettait de déclencher un chronomètre et l'enregistrement des données comportementales sous forme de rapport d'activité.

La tâche des participants consistait à apprendre une procédure d'assemblage du cube de Soma, un puzzle 3D monochrome couleur bois constitué de sept pièces de formes différentes. Une modélisation des pièces du cube était présente au sein de l'EV développé (Fig. 1). Des instructions graphiques étaient disponibles sur un panneau divisé en plusieurs cases correspondant chacune à une étape de la procédure. Afin de consulter une instruction, le participant devait passer la manette sur la case souhaitée pour rendre visible le schéma de montage. Enfin, une grille d'assemblage marquait l'emplacement de construction du cube.

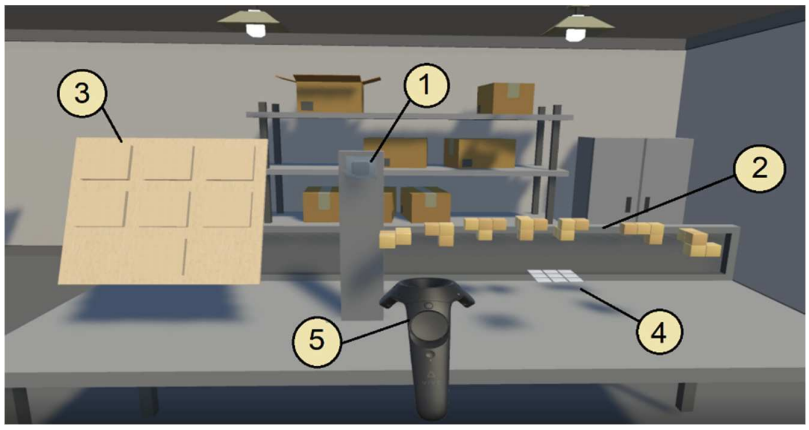


Fig. 1. Vue de l'EVAH. (1) Interrupteur. (2) Les sept pièces du cube de Soma. (3) Panneau d'instructions. (4) Grille d'assemblage. (5) Représentation virtuelle de la manette.

3.3 Matériel

L'immersion dans l'EV était réalisée grâce à un casque de RV HTC Vive Pro Eye (2019). Les participants étaient également équipés d'un contrôleur Vive dans leur main dominante pour interagir avec l'EV. Une représentation de la manette Vive dans l'EV était présente mais les mains du participant n'étaient pas modélisées (Fig. 1). Les participants avaient seulement besoin d'utiliser la gâchette à l'arrière de la manette Vive pour effectuer la tâche procédurale.

Un ordinateur portable avec les logiciels LimeSurvey et Tholos [20] permettait d'afficher différents questionnaires : questionnaire sociodémographique, SSQ [19] et NASA-TLX [21]. Une grille d'observation, version papier, a également été utilisée lors de chaque passation par l'expérimentatrice.

3.4 Données recueillies

Afin d'examiner le déroulement de l'apprentissage de procédure, plusieurs mesures basées sur les travaux existants dans le domaine de l'apprentissage de procédure ont été recueillis [3 ; 4]. Les mesures comportementales impliquaient des mesures chronométriques et des mesures non chronométriques. Elles correspondaient au temps de manipulation des pièces du cube (TM), au temps de consultation des instructions (TC), au temps d'inactivité ou de latence (TL), au temps total de réalisation de la tâche (TRT, i.e., somme des TM, TC et TL), au nombre de manipulations des pièces, au nombre de consultations des instructions et au nombre d'erreurs ou d'actions incorrectes (i.e., action qui ne correspond pas à l'instruction en cours en terme de pièce et/ou de position attendue). Le nombre d'essais était également recueilli. Ces mesures étaient complétées par une mesure subjective de la charge de travail obtenue à l'aide de l'échelle NASA-TLX [21].

3.5 Procédure

L'expérience a duré au total environ 1h30 pour chaque participant et était divisée en deux sessions individuelles espacées d'une semaine (plus ou moins un jour). Chaque session se terminait lorsque le participant atteignait un critère d'arrêt défini au préalable (ou un temps de passation supérieur à 1h00) afin d'amener chaque participant à une performance similaire. Le critère était de réaliser trois fois et de façon successive la procédure d'assemblage sans erreur et sans consultation d'instructions.

La première session, d'une durée d'environ 1h00, était subdivisée en cinq phases. (i) Le briefing : les participants devaient remplir un formulaire de consentement, un questionnaire sociodémographique et le SSQ. (ii) La démonstration : une présentation de l'EVAH, de la tâche et des interactions à accomplir avec le contrôleur Vive était réalisée par l'expérimentatrice via des consignes formelles. (iii) La familiarisation : une procédure d'assemblage de pièces géométriques colorées a été conçue pour cette phase. Il ne s'agissait pas des pièces du cube de Soma afin d'éviter des biais d'apprentissage. Les participants, équipés du casque de RV exécutaient deux essais successifs de la tâche procédurale afin de se familiariser avec le matériel et avec une procédure à exécuter dans un ordre précis. (iv) L'apprentissage : le participant équipé du casque de RV, devait exécuter plusieurs fois la procédure d'assemblage du cube de Soma jusqu'à l'atteinte du critère d'arrêt. Entre chaque essai, le questionnaire NASA-TLX était présenté au participant, qui devait alors enlever le casque de RV. (v) Le débriefing : le post-questionnaire SSQ était administré au participant et était suivi d'un moment de discussion avec l'expérimentatrice.

La deuxième session, réalisée après un délai d'une semaine, durait environ 30 minutes et était effectuée pour vérifier l'acquisition de la procédure. Elle comportait une phase de briefing et de débriefing, avec une phase d'évaluation similaire à la phase d'apprentissage de la session 1.

4 Résultats

Les analyses statistiques ont été effectuées à l'aide du logiciel R avec les *packages* « lmerTest », « multcomp » et « FactoMineR » [22 ; 24]. La méthode d'approximation de Satterthwaite a été utilisée pour le calcul des degrés de liberté du dénominateur ainsi que de la statistique F. Si nécessaire, des tests post-hoc par paire (avec ajustement de Bonferroni) ont été effectués. Trois individus, aux données extrêmes, ont été exclus des analyses. Les résultats présentés ci-après concernent les données de la session 1. Les analyses de la session 2 montrent que l'ensemble des indicateurs comportementaux étaient proches des résultats obtenus à la fin de la première session indiquant qu'une récupération de la procédure dans la mémoire à long terme est possible après une interruption d'une semaine malgré quelques difficultés lors de la première exécution. La Fig. 2 présente l'évolution des mesures comportementales chronométriques lors de l'apprentissage procédural, et la Fig. 3 présente les classes obtenues à l'issue de l'application de la méthode de *clustering* sur les données comportementales.

4.1 Évaluation de l'apprentissage procédural cognitif

Les participants réalisent en moyenne 6 essais lors de la session 1 pour atteindre le critère d'arrêt (σ = 1,32 ; *min* = 4 ; *max* = 9).

Données comportementales. Les mesures chronométriques liées à la réalisation de la tâche présentent l'allure classique d'une courbe d'apprentissage (Fig. 2). Par ailleurs, la variabilité interindividuelle, élevée au premier essai, diminue également au fil des essais, marquant l'effet de l'apprentissage [3].

Le Modèle Linéaire Mixte (MLM) confirme l'effet du nombre d'essais sur les variations des mesures comportementales chronométriques : pour le temps total de réalisation de la tâche : F (18 ; 633,42) = 51,03 ; p < .001, η^2_p = 0,59, pour le temps d'inactivité : F (18 ; 632,02) = 45,57 ; p < .001, η^2_p = 0,56, pour le temps de manipulation des pièces : F (18 ; 630,95) = 15,96 ; p < .001, η^2_p = 0,31 et pour le temps de consultation des instructions : F (18 ; 641,64) = 60,74 ; p < .001, η^2_p = 0,63. Les tests post-hoc montrent qu'au fur et à mesure des essais, les participants réalisent plus rapidement la tâche, avec un temps d'inactivité, de manipulation des pièces et de consultation des instructions moindre entre les essais 1 et 2 de la session 1 (p < .001). L'analyse montre également une différence significative entre l'essai 2 et l'essai 3 pour le temps d'inactivité (p < .01).

Les mesures comportementales non chronométriques recueillies diminuent également fortement au fur et à mesure des répétitions de la tâche. Le modèle linéaire mixte confirme l'effet du nombre d'essais sur les variations des mesures non chronométriques : pour le nombre de manipulations des pièces : F (18 ; 630,21) = 15,05 ; p < .001, η^2_p = 0,30, pour le nombre de consultations des instructions : F (18 ; 641,2) = 65,51 ; p < .001, η^2_p = 0,65, et pour le nombre d'erreurs : F (18 ; 642,62) = 14,96 ; p < .001, η^2_p = 0,30. Les tests post-hoc montrent qu'entre les essais 1 et 2 de la session 1, les participants réalisent la procédure avec moins de consultations d'instructions, une manipulation des pièces plus précise et commettent moins d'erreurs (p < .001). Les

analyses révèlent également une diminution du nombre de consultations des instructions entre l'essai 2 et l'essai 3 ($p < .001$). Ces résultats confirment H1 puisqu'une amélioration puis une stabilisation des performances sont observées au fur et à mesure des répétitions de la tâche en session 1.

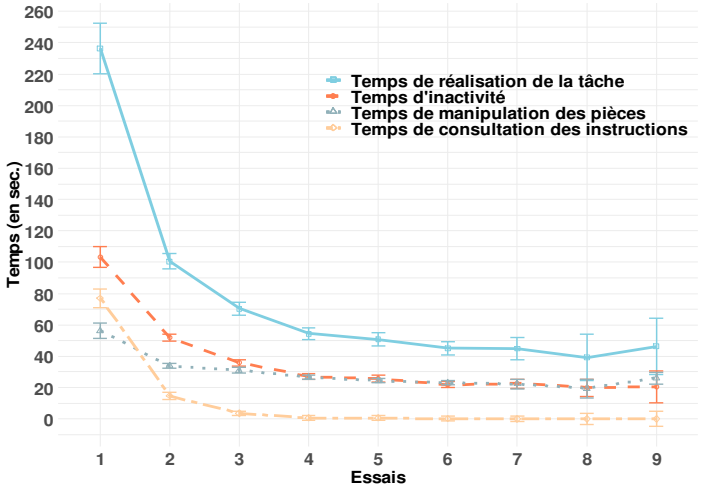


Fig. 2. Mesures chronométriques moyennes (en sec.) lors des essais (barres d'erreurs avec l'erreur standard de la moyenne calculée pour les modèles intra-participants [25]).

Données subjectives. Un score de charge de travail mental multidimensionnel a été calculé sans prendre en compte le critère d'exigence physique.

Les scores obtenus à l'échelle du NASA-TLX évoluent au fur et à mesure des essais. Le modèle linéaire mixte confirme l'effet du nombre d'essais sur le score de charge de travail mental calculé à partir de cinq critères : F (18 ; 625,07) = 26,77 ; $p < .001$, η^2_p = 0,44. Les tests post-hoc montrent qu'au fil des répétitions de l'exécution de la procédure, l'indice de charge de travail mental évolue entre certains essais successifs. Les analyses montrent une diminution significative entre l'essai 1 et l'essai 2 ($p < .01$) et entre l'essai 2 et l'essai 3 ($p < .05$). Ces résultats confirment H2 puisque le niveau de charge de travail mental ressenti est élevé au début de l'apprentissage puis diminue jusqu'à aboutir à une absence de fluctuations en session 1.

4.2 Identification de phases d'apprentissage

Pour identifier différentes phases d'apprentissage, une méthode de *clustering* non supervisée nommée Classification Ascendante Hiérarchique (CAH) a été réalisée sur les composantes principales d'une analyse factorielle (ici une ACP, Analyse en Composantes Principales). Une étape de consolidation par *k-means* a été effectuée après le découpage de l'arbre afin de minimiser la variance intra-classe. Cette méthode présente l'avantage de ne pas nécessiter la détermination préalable d'un nombre de classes, une coupure étant proposée par défaut selon la partition ayant la plus grande perte relative

d'inertie. Le jeu de données utilisé comprend six données comportementales de la session 1 (pour une absence de redondance dans les variables, le TRT a été écarté) obtenues pour chaque individu à chaque essai. La CAH est effectuée sur les coordonnées des individus sur les 2 premiers axes factoriels de l'ACP normée. Les paramètres utilisés sont le critère de Ward et les distances euclidiennes. Les individus et les classes sont regroupés en fonction de l'inertie.

Une visualisation des classes obtenues est présentée à la Fig. 3. Les variables qui expliquent le mieux un découpage en trois classes sont le nombre de consultations des instructions ($\eta^2 = 0{,}74$), le temps de consultation des instructions ($\eta^2 = 0{,}73$) et le temps d'inactivité ($\eta^2 = 0{,}72$). La classe 1 compte 320 données, tandis que la classe 2 comprend 58 données et la classe 3, seulement 11 données.

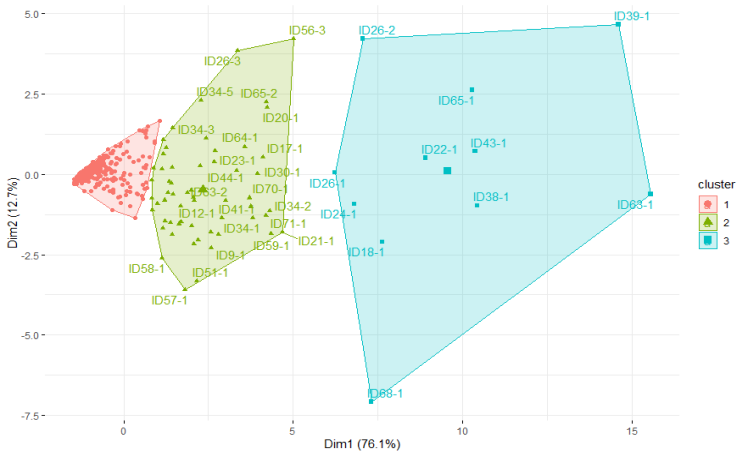


Fig. 3. Visualisation des *clusters* obtenus par CAH sur une ACP.

Les classes 1 et 2 se distinguent par des performances opposées puisque la classe 2 comprend l'activité d'apprenants caractérisée par un nombre et un temps de consultation des instructions ainsi qu'un temps d'inactivité important comparativement aux données de la classe 1 caractérisées par de meilleures performances (Tableau 1). Contrairement à ces classes, la variable qui explique le mieux la classe 3 est le nombre d'erreurs commises qui est relativement élevé (Tableau 1). Cette classe semble représenter des individus aux comportements « extrêmes » dans le jeu de données (Tableau 1). Plus précisément, la classe 2 est représentée à environ 66% par le premier essai des participants alors que la classe 1 comprend l'ensemble des performances obtenues lors des essais 6 à 9. Une analyse par MLM montre un effet de la classe sur le score de charge de travail mental : $F\ (2\ ;\ 338{,}44) = 103{,}94$; $p < .001$, $\eta^2_p = 0{,}38$. Les tests post-hoc révèlent des différences entre chaque classe avec un score plus élevé pour la classe 3 ($\mu = 61{,}2$; $\sigma = 12{,}8$), puis pour la classe 2 ($\mu = 41{,}8$; $\sigma = 16{,}9$), et enfin le score le plus faible pour la classe 1 ($\mu = 22{,}8$; $\sigma = 15{,}9$) ($p < .001$). Ces résultats confirment H3 avec l'identification possible de deux phases lors du processus d'apprentissage qui correspondraient pour les classes 2 et 3 au début de l'apprentissage et pour la classe 1 au

milieu et à la fin de l'apprentissage. Toutefois, contrairement aux théories et aux travaux présentés, cette approche de *clustering* ne permet pas le repérage des trois phases d'apprentissage et semble sensible aux différences interindividuelles avec l'identification de la classe 3.

Tableau 1. Présentation de caractéristiques des classes obtenues par CAH sur une ACP (moyenne de chaque classe fournie et écart-type donné entre parenthèses).

Mesures comportementales	**Classe 1**	**Classe 2**	**Classe 3**
Nombre de consultations des instructions	2,41 (*5,22*)	30,81 (*17,11*)	80,27 (*28,35*)
Temps de consultation des instructions (en sec.)	2,84 (*7,19*)	55,36 (*36,68*)	152,53 (*59,20*)
Temps d'inactivité (en sec.)	28,53 (*14,96*)	89,33 (*26,32*)	203,10 (*81,06*)
Nombre d'erreurs	0,81 (*2,04*)	7,83 (*7,26*)	35,09 (*17,73*)

5 Conclusion

Les résultats montrent qu'une exécution répétée de la tâche en EV induit de meilleures performances. La charge de travail mental diminue également au fil des répétitions suivant des profils de résultats similaires à l'évolution comportementale.

Après avoir constaté qu'il y avait bien eu un apprentissage de la procédure en EV avec la possibilité de récupérer la procédure en mémoire à long terme, l'objectif était d'identifier des phases d'apprentissage à partir de l'application d'une méthode de *clustering* sur un ensemble de données comportementales. L'interprétation des classes obtenues permet d'identifier des phases de l'apprentissage procédural. En effet, deux classes s'opposent à la fois au niveau des caractéristiques comportementales, des essais et des scores de charge de travail mental. Comme Anderson [10] et Ganier et al. [3] l'ont évoqué, les phases d'apprentissage se distinguent par les processus cognitifs en jeu induisant des comportements et un coût cognitif de traitement de l'information différents. La première phase, appelée cognitive, correspond à la découverte de la procédure caractérisée par de nombreuses erreurs, une nécessité de consulter les instructions longtemps et de façon répétée et une charge de travail mental élevée. Ceci correspondrait aux classes 2 et 3 marquées également par un temps d'inactivité important pouvant refléter le profil novice des apprenants avec un temps de réflexion sur la compréhension des instructions ou du matériel. La deuxième phase, nommée associative, serait illustrée par une exécution de la procédure qui devient plus fluide et rapide avec une détection progressive des erreurs et un besoin moins marqué des instructions. La troisième phase, intitulée autonome, se distingue quant à elle, par une performance optimale avec une absence d'erreurs et de consultations d'instructions. Ainsi, la classe 1 pourrait faire référence aux phases associative et autonome puisqu'elle se distingue des autres classes par de meilleures performances et une faible charge de travail mental. Toutefois, cette méthode de *clustering* ne permet pas de repérer trois phases distinctes lors du processus d'apprentissage comme proposées par les théories et distinguées par les travaux précé-

dents. Cette difficulté rejoint les propos de Tenison et Anderson [4] concernant la complexité de distinguer la phase 2 et 3 d'un point de vue comportemental. Par ailleurs, la CAH semble sensible aux différences de profils d'apprentissage ce qui affecterait l'identification des phases d'apprentissage. Cette méthode présente donc quelques défauts et pourrait être améliorée par exemple par une étape de regroupement des apprenants selon leur profil ou par l'ajout d'une donnée de progression entre essais. Il s'agit également d'une solution d'identification *a posteriori* mais qui pourrait permettre de labelliser des données afin de développer une approche d'apprentissage supervisé et ainsi pouvoir individualiser en temps réel et de façon automatique les scénarios pédagogiques selon la phase d'apprentissage atteinte.

Des travaux futurs sont envisagés à partir d'une approche de distinction des phases d'apprentissage s'appuyant sur les propositions d'Hubert et al. [12] et de Ganier et al. [3]. Contrairement à la démarche d'identification des phases employée dans cette étude, il s'agirait d'émettre des règles comportementales basées sur la théorie ACT-R [13]. Ces deux démarches pourraient être complémentaires dans la recherche de critères comportementaux permettant d'identifier le passage d'une phase à l'autre.

Remerciements et Financements

Ce travail a été soutenu par un financement du gouvernement français géré par l'ANR, numéro de subvention ANR-21- ESRE-0030 (CONTINUUM) et également grâce au soutien de Brest Métropole.

References

1. Rodriguez, J., Gutirrez, T., J., E., Casado, S., Aguinag, I.: Training of Procedural Tasks Through the Use of Virtual Reality and Direct Aids. In: Sk Lnyi, C. (ed.) Virtual Reality and Environments. InTech (2012). https://doi.org/10.5772/36650.
2. Spittle, M.: Motor Learning and Skill Acquisition: Applications for Physical Education and Sport. Bloomsbury Publishing, Londres (2021).
3. Ganier, F., Hoareau, C., Devillers, F.: Évaluation des performances et de la charge de travail induits par l'apprentissage de procédures de maintenance en environnement virtuel. Trav. Hum. 76, 335–363 (2013). https://doi.org/10.3917/th.764.0335.
4. Tenison, C., Anderson, J.R.: Modeling the distinct phases of skill acquisition. J. Exp. Psychol. Learn. Mem. Cogn. 42, 749–767 (2016). https://doi.org/10.1037/xlm0000204.
5. Siemens, G.: Learning analytics: envisioning a research discipline and a domain of practice. In: Proceedings of the 2nd International Conference on Learning Analytics and Knowledge. pp. 4–8. ACM, Vancouver British Columbia Canada (2012). https://doi.org/10.1145/2330601.2330605.
6. Sun, P.-C., Tsai, R.J., Finger, G., Chen, Y.-Y., Yeh, D.: What drives a successful e-Learning? An empirical investigation of the critical factors influencing learner satisfaction. Comput. Educ. 50, 1183–1202 (2008). https://doi.org/10.1016/j.compedu.2006.11.007.
7. Deeva, G., Bogdanova, D., Serral, E., Snoeck, M., De Weerdt, J.: A review of automated feedback systems for learners: Classification framework, challenges and opportunities. Comput. Educ. 162, (2021). https://doi.org/10.1016/j.compedu.2020.104094.

8. Anderson, D.I., Lohse, K.R., Lopes, T.C.V., Williams, A.M.: Individual differences in motor skill learning: Past, present and future. Hum. Mov. Sci. 78, (2021). https://doi.org/10.1016/j.humov.2021.102818.
9. Heurley, L., Ganier, F.: L'utilisation des textes procéduraux : lecture, compréhension et exécution d'instructions écrites. Intellectica. 44, 45–62 (2006). https://doi.org/10.3406/intel.2006.1291.
10. Anderson, J.R.: Acquisition of cognitive skill. Psychol. Rev. 89, 369–406 (1982).
11. Fitts, P.M., Posner, M.I.: Human Performance. Brooks/Cole Publishing, Belmont, CA (1967).
12. Hubert, V., Beaunieux, H., Chételat, G., Platel, H., Landeau, B., Danion, J.-M., Viader, F., Desgranges, B., Francis Eustache: The dynamic network subserving the three phases of cognitive procedural learning. Hum. Brain Mapp. 28, 1415–1429 (2007). https://doi.org/10.1002/hbm.20354.
13. Corbett, A.T., Anderson, J.R.: Knowledge tracing: Modeling the acquisition of procedural knowledge. User Model. User-Adapt. Interact. 4, 253–278 (1994). https://doi.org/10.1007/BF01099821.
14. Rasmussen, J.: Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models. IEEE Trans. Syst. Man Cybern. SMC-13, 257–266 (1983). https://doi.org/10.1109/TSMC.1983.6313160.
15. VanLehn, K.: Cognitive Skill Acquisition. Annu. Rev. Psychol. 47, 513–539 (1996). https://doi.org/10.1146/annurev.psych.47.1.513.
16. Taylor, J.A., Ivry, R.B.: The role of strategies in motor learning. Ann. N. Y. Acad. Sci. 1251, 1–12 (2012). https://doi.org/10.1111/j.1749-6632.2011.06430.x.
17. Evans, N.J., Brown, S.D., Mewhort, D.J.K., Heathcote, A.: Refining the law of practice. Psychol. Rev. 125, 592–605 (2018). https://doi.org/10.1037/rev0000105.
18. Tenison, C., Fincham, J.M., Anderson, J.R.: Phases of learning: How skill acquisition impacts cognitive processing. Cognit. Psychol. 87, 1–28 (2016).
19. Kennedy, R.S., Lane, N.E., Berbaum, K.S., Lilienthal, M.G.: Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness. Int. J. Aviat. Psychol. 3, 203–220 (1993). https://doi.org/10.1207/s15327108ijap0303_3.
20. Cegarra, J., Chevalier, A.: The use of Tholos software for combining measures of mental workload: Toward theoretical and methodological improvements. Behav. Res. Methods. 40, 988–1000 (2008).
21. Hart, S., Staveland, L.: Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. Adv. Psychol. 52, 139–183 (1988). https://doi.org/10.1016/S0166-4115(08)62386-9.
22. Bretz, F., Hothorn, T., Westfall, P.: Multiple comparisons using R. Chapman and Hall/CRC, Londres (2016).
23. Kunzetsova, A., Brockhoff, P., Christensen, R.: lmerTest package: Tests in linear mixed effect models. J Stat Softw. 82, 1–26 (2017).
24. Lê, S., Josse, J., Husson, F.: FactoMineR: an R package for multivariate analysis. J. Stat. Softw. 25, 1–18 (2008).
25. Cousineau, D.: Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. Tutor. Quant. Methods Psychol. 1, 42–45 (2005).

Conception d'un système de révisions : une étude qualitative exploratoire pour identifier les besoins des enseignants et des élèves

Denis Ollivier[1,3], Franck Silvestre[1,2][0000−0002−1134−8200], Jean-Baptiste Raclet[2][0000−0001−7357−912X], Emmanuel Lescure[3] et Julien Broisin[2][0000−0001−8713−6282]

[1] IRIT, Université Toulouse Capitole
[2] IRIT, Université Toulouse III - Paul Sabatier
[3] Kosmos Education

Résumé. Pour aider les élèves dans l'activité de révisions, concevoir un outil exploitant les effets test (se tester au lieu de relire), et d'espacement (espacer les révisions dans le temps) semble pertinent mais peu d'études en contextes écologiques en font l'objet. Dans le cadre d'une démarche orientée par la conception, ce papier présente l'étude qualitative exploratoire menée auprès d'enseignants et d'élèves du secondaire visant à répondre aux questions de recherches suivantes : Quelles hypothèses aideront à concevoir un outil de révisions adapté (QR1) aux besoins des enseignants et (QR2) aux besoins des élèves ? Les échanges ont permis d'identifier plusieurs hypothèses, telles que (H2) l'outil aidera les élèves à se tester plus,(H3) même si les révisions resteront massées autour des examens. Ces hypothèses seront testées à travers une étude quantitative, dont les données seront collectées avec une première version du système de révision, qui sera améliorée à travers les prochaines itérations.

Mots-clés : Effet test , Effet d'espacement , Révisions , Système interactif , Méthode mixte

Abstract. To help students in the revision activity, designing a tool exploiting the test (testing oneself instead of rereading) and spacing effects (spacing revisions in time) seems relevant but few studies in ecological contexts have done so. Within the framework of a design-based research, this paper presents the exploratory qualitative study conducted with secondary school teachers and students to answer the following research questions: What assumptions will help to design a revision tool adapted to (RQ1) teachers' needs and (RQ2) students' needs? The discussions identified several hypotheses, such as (H2) the tool will help students to test themselves more,(H3) even if the revisions will remain massed around the exams. These hypotheses will be tested through a quantitative study, whose data will be collected with a first version of the revision system, which will be improved through the next iterations.

Keywords. Testing effect , Spacing effect , Revision , Interactive system , Mixed method

1 Révisions et numérique

L'apprentissage des élèves ne s'arrête pas aux enseignements reçus en classe. Comme le mentionne le programme de l'éducation nationale française dans son bulletin officiel N°27 du 02 Juillet 2015, l'un des objectifs des enseignements est d'aider les élèves à acquérir les méthodes nécessaires à l'apprentissage, telles que les révisions. Pour travailler autour de cette activité de révision nous proposons la définition suivante dérivée de celle fournie par le Larousse en ligne[4] : "l'action d'étudier de nouveau une ressource pédagogique (leçon, exercice, etc.) en vue d'une évaluation sommative".

Différentes études ont montré que réviser une connaissance est plus efficace en se testant sur cette connaissance qu'en relisant les leçons qui la présentent [10, 12]. Par ailleurs, l'espacement des séances de travail [6, 11] est plus efficace qu'un effort massé sur une seule séance. Ces effets connus respectivement sous le nom d'effet test et d'espacement sont tous les deux bénéfiques pour la rétention à long terme des informations.

La littérature questionne déjà le soutien de ces effets via l'outil numérique, comme l'a fait Latimier dans son étude expérimentale avec l'outil Didask, conçus pour l'apprentissage de connaissances déclaratives [8]. Dans ce cadre, Latimier précise néanmoins dans les pages 144 à 147 que les conditions expérimentales ne correspondent pas à un contexte écologique, puisque les élèves étaient des adultes en parfaite autonomie, essentiellement motivés par une compensation financière, et sans interaction avec un professeur, ou d'autres élèves. Latimier [8] souligne alors que les populations plus jeunes sont rarement testées dans les études sur les effets test et d'espacement et qu'il serait intéressant de le faire.

Nos travaux de recherche s'intéressent donc à la conception d'un système interactif de révisions tirant parti des effets test et d'espacement dans le secondaire, en conditions scolaires et s'inscrivent dans l'alignement constructiviste de Biggs [4]. Afin d'aligner les objectifs pédagogiques avec l'activité d'apprentissage et l'activité d'évaluation, les révisions qui visent à préparer une évaluation sommative doivent s'appuyer sur des activités menées en classes évaluées de manière formative. C'est pourquoi nous avons décidé de construire un outil qui s'appuie sur Elaastic[5], anciennement connu sous le nom de *Tsaap-Notes*. Cet outil a déjà fait l'objet d'études en contextes écologiques dans l'enseignement supérieur [2] et dans l'enseignement secondaire [1] mais elles n'ont pas exploré dans le détail les modalités de révisions, comme l'espacement des tests de révision par exemple.

Cet article présente l'étude qualitative exploratoire menée auprès d'enseignants et d'élèves avec les questions de recherche suivantes :

QR1. Quelles hypothèses aideront à concevoir un outil de révisions adapté aux besoins des enseignants ?
QR2. Quelles hypothèses aideront à concevoir un outil de révisions adapté aux besoins des élèves ?

[4] https://www.larousse.fr/dictionnaires/francais/révision/69135
[5] https://www.irit.fr/elaastic/

2 L'étude qualitative

Méthode de conduite de la recherche Afin de concevoir un outil adapté aux besoins des enseignants et des élèves, nous souhaitons suivre une conception centrée humain. Notre travail s'inscrit dans une démarche de recherche orientée par la conception [9], et nous abordons nos itérations à travers des méthodes mixtes pour capitaliser sur les apports des approches qualitatives et quantitatives.

Pour la première itération de ce projet de recherche nous suivons le modèle séquentiel exploratoire de Creswell et Clark [7]. En accord avec ce modèle, nous avons réalisé une étude qualitative auprès d'enseignants et d'élèves volontaires, afin d'identifier les hypothèses qui aideront à la conception d'un système interactif de révision adapté aux besoins des enseignants et des élèves.

Participants Nous avons interrogés 8 enseignants déjà formés à l'utilisation d'Elaastic. Seuls 6 enregistrements valides ont pu être conservés. En effet, un enregistrement endommagé, ainsi qu'un entretien pour lequel l'autorisation écrite d'utilisation des données n'a pas été donnée, ont dû être exclus.

Dans un second temps, nous avons pu obtenir 4 entretiens avec des élèves ayant récemment terminé leurs études dans le secondaire, ou en dernière année.

Procédure de collecte de données L'étude qualitative a été réalisée à partir d'entretiens individuels semi-dirigés d'environ 45 minutes qui s'appuyaient sur un questionnaire différent selon qu'il s'agisse des enseignants ou des élèves. Les questionnaires ont été en partie construits en s'appuyant sur l'UTAUT2.

Les entretiens étaient découpés en 3 phases. La première phase visait une meilleure compréhension de l'environnement de travail des participants. La deuxième phase était la présentation d'un prototype d'outil de révisions afin de rendre le sujet concret pour les participants. Enfin la dernière phase consistait à interroger les participants sur le prototype présenté. Une analyse par encodage a été effectuée sur les entretiens pour mettre en avant les thèmes les plus pertinents. Dans une démarche de science ouverte, les questionnaires, les retranscriptions ainsi que les données recueillies sont accessibles à ce lien [6]

Le prototype présenté lors des entretiens était constitué d'écrans statiques non fonctionnels, mais d'apparence haute-fidélité avec une approche *mobile first*. Il a été conçu avec l'idée de proposer un contenu déjà travaillé en classe à travers l'utilisation de l'outil en ligne Elaastic [2].

3 Résultats des entretiens

Adapter l'outil à tous les équipements. Plusieurs enseignants ont évoqué le besoin que l'outil soit adapté à différents équipements, certains en défaveur du mobile : En1-"[...] ils ont pas de portables ils ont parfois une tablette ou quand même un ordi à la maison [...]", d'autres en faveur de celui-ci : En4-"[...] les élèves savent

[6] https://osf.io/cz7fe/?view_only=cc64c8417d184a9298efc638a54cde34

moins utiliser l'ordinateur bizarrement aujourd'hui que la tablette ou le téléphone". D'autre part, les élèves encouragent aussi cette idée, certains préférant l'ordinateur lorsqu'il s'agit de travailler : El3-"[...] Moi en tant qu'étudiant j'utilise principalement l'ordinateur, après les deux plateformes sont bonnes je pense.".

L'engouement lié à l'usage du numérique. Les enseignants font le constat que le numérique motive intrinsèquement les élèves ce qui les motive aussi : En2-"Tous mes cours sont systématiquement illustrés par un diaporama [...] Je sais qu'ils l'apprécient beaucoup [...] ", En1-"... je sais qu'ils aiment bien quand il y a des petites vidéos ou des choses en plus [...] ". de leur côté, les élèves ont réagi avec entrain aux maquettes présentées, illustrant directement l'engouement dont ils font part à l'égard du numérique : El2-"Alors moi je trouve ça très intéressant, surtout le fait que du coup les élèves puissent progresser en voyant l'argumentation de leur camarades vraiment je trouve ça très intéressant.'.

Suivre les élèves en encadrant les révisions Interroger les enseignants sur des indicateurs du travail des élèves fait ressortir leur intérêt sur de telles informations dans le but d'accompagner les élèves : En5-"Alors ce qui pourrait être intéressant c'est de savoir le taux d'utilisation de l'application [...] pour les aider à mieux répartir leur charge de travail.". Enfin d'autres précisent qu'un tel accompagnement peut induire une charge de travail qu'il faut considérer : En8-"Le temps ou la charge de travail je pense il est vraiment sur l'analyse du retour [...]"Cet accompagnement résonne avec l'avis exprimé par les élèves sur le rôle de l'enseignant : El2-"Au lycée [...] on se réfère beaucoup aux professeurs [...] donc je pense qu'inciter c'est le plus important".

Une éventuelle notation des révisions Quelques enseignants soulignent qu'une notation pourrait être nécessaire en tant que motivation extrinsèque : En4-"[...] peut-être va falloir justement trouver une carotte pour justement donner l'envie d'y aller spontanément [...]". Cependant les élèves semblent penser qu'encadrer par la notation pourrait décourager : El4-"[...] une notation peut démoraliser l'élève en fonction des caractères, et il peut se dire « Je suis bloqué à telle note » et ça peut braquer surtout.".

4 Discussion

Un système adapté au élèves (QR2) Tout d'abord, pour bénéficier des effets test et espacements via l'outil, il faut s'assurer que les élèves puissent l'utiliser. Nous pensions d'abord à tort que proposer un accès via mobile, source de motivation pour les élèves [3], pouvait suffire. Nous avons pu réviser notre jugement suite aux entretiens, et supposons maintenant qu'il faut fournir la possibilité d'accéder à l'outil via différents équipements et observer à travers lesquels les élèves accéderont à l'outil (H1).

Les entretiens ont renforcé l'idée qu'utiliser un outil numérique tel que le nôtre motive les élèves. Ainsi nous supposons que mettre à disposition un outil

mettant en place l'effet test permettra d'observer un effet bénéfique sur la pratique du test comme méthode de révisions (H2).

Toutefois, certains enseignants soulignent que cette motivation ne sera peut-être pas suffisante en l'absence de notation. Ainsi, l'absence de notation autour de cette activité peut laisser penser qu'un manque d'activité sera observable, à l'exception des périodes d'examens autour desquelles les élèves auront une motivation pour réviser. D'ailleurs, certaines études sur le comportement des élèves vis-à-vis des révisions montrent une tendance des élèves à réviser en masse à l'approche des examens [5]. On peut donc émettre l'hypothèse que l'usage du système de révision sera marquée à l'approche des examens si les élèves ne sont pas sollicités explicitement par l'outil (H3).

Un système adapté aux enseignants (QR1) D'autre part, les enseignants ont réagi positivement à l'idée d'avoir des indicateurs les informant sur la façon de réviser des élèves. Cependant, un besoin de vigilance quant au temps que ce suivi peut coûter a été mis en avant. Nous supposons donc que les enseignants souhaitent avoir accès à des données nominatives sur le travail des élèves afin d'améliorer le suivi individualisé malgré la charge de travail induite (H4).

Enfin, une telle collecte de données attire notre attention sur la vie privée des étudiants et les données à caractère personnelles. Nous allons concevoir l'outil en conséquence afin qu'il garantisse la transparence sur l'utilisation des données personnelles, et offre un contrôle sur la collecte, même si nous supposons que les élèves et leurs parents accepteront l'usage des données à des fins pédagogiques (H5).

Limites de l'étude Dans notre étude, les profils des participants d'une même catégorie sont plutôt similaires, et les élèves ont moins développé leurs idées sur les questions ouvertes. Ces limites inhérentes à l'étude qualitative seront palliées par l'étude quantitative qui suit naturellement dans le cadre de méthodes mixtes.

Table 1. Les hypothèses identifiées à l'aide des entretiens

	Hypothèses
H1	Les élèves vont accéder au système de révisions avec divers équipements
H2	Fournir aux élèves un outil de révisions mettant en place l'effet test aura un effet bénéfique sur leur pratique du test comme méthode de révisions
H3	L'usage de l'outil sera particulièrement marqué à proximité des examens
H4	Les enseignants souhaitent avoir accès à des données nominatives pour améliorer le suivi individualisé malgré la charge de travail induite
H5	Les élèves et leurs parents accepteront que l'on collecte des données à caractère personnel pour construire des indicateurs pour les enseignants

5 Conclusion

Dans le cadre de nos recherches, nous nous demandons comment concevoir un système interactif adapté à l'activité de révisions, en bénéficiant des effets test et d'espacement. Nous avons décidé d'interroger des enseignants et des élèves pour identifier les hypothèses qui seront vérifiées dans l'étude quantitative. Ces hypothèses sont réunies dans le **tableau 1**. Nous allons développer une première version de l'outil permettant de vérifier ces hypothèses, et chercherons à identifier des recommandations pour la conception d'outils de révisions adaptés aux besoins des élèves et des enseignants à travers les prochaines itérations de notre recherche orientée par la conception.

References

1. Andriamiseza, R.: Learning Analytics-Based Formative Assessment Recommendations for Technology-Enhanced Learning Practices. Ph.D. thesis, Université de Toulouse, Université Toulouse III-Paul Sabatier (2022)
2. Andriamiseza, R., Silvestre, F., Parmentier, J.F., Broisin, J.: Recommendations for orchestration of formative assessment sequences: a data-driven approach. In: European Conference on Technology Enhanced Learning. pp. 245–259. Springer (2021)
3. Bai, H.: Pedagogical practices of mobile learning in k-12 and higher education settings. TechTrends **63**(5), 611–620 (2019)
4. Biggs, J.: Enhancing teaching through constructive alignment. Higher education **32**(3), 347–364 (1996)
5. Blasiman, R.N., Dunlosky, J., Rawson, K.A.: The what, how much, and when of study strategies: Comparing intended versus actual study behaviour. Memory **25**(6), 784–792 (2017)
6. Carpenter, S.K., Cepeda, N.J., Rohrer, D., Kang, S.H., Pashler, H.: Using spacing to enhance diverse forms of learning: Review of recent research and implications for instruction. Educational Psychology Review **24**(3), 369–378 (2012)
7. Creswell, J.W., Plano Clark, V.L., Gutmann, M.L., Hanson, W.E.: Advanced mixed methods research designs. Handbook of mixed methods in social and behavioral research **209**(240), 209–240 (2003)
8. Latimier, A.: Optimisation de l'apprentissage par récupération en mémoire pour promouvoir la rétention à long terme de nouvelles connaissances. Ph.D. thesis, Université Paris sciences et lettres (2019)
9. 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)
10. Phelps, R.P.: The effect of testing on student achievement, 1910–2010. International Journal of Testing **12**(1), 21–43 (2012)
11. Rohrer, D., Taylor, K.: The effects of overlearning and distributed practise on the retention of mathematics knowledge. Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition **20**(9), 1209–1224 (2006)
12. Rowland, C.A.: The effect of testing versus restudy on retention: a meta-analytic review of the testing effect. Psychological bulletin **140**(6), 1432 (2014)

Étudier l'expérience des visiteurs pour identifier leurs processus d'apprentissage au musée

Vers une conception d'EIAH adaptés à la visite non guidée

Marine Lagasse[1], Daniel Schmitt[2] et Bruno De Lièvre[3]

[1] Université de Mons, IPN & Université Polytechnique Hauts-de-France, LARSH
marine.lagasse@umons.ac.be
[2] Université Polytechnique Hauts-de-France, LARSH, 59313 Valenciennes, France
daniel.schmitt@uphf.fr
[3] Université de Mons, IPN, 7000 Mons, Belgique
bruno.delievre@umons.ac.be

Résumé. Notre étude de l'expérience de visite se base sur les entretiens en rappel stimulé de 52 visiteurs de musées français et belges. L'accès aux micro-dynamiques de leur expérience permet de connaître leurs processus cognitifs et sensoriels adoptés *in situ*. Leur analyse éclaire finalement les processus d'apprentissage favorisés par le contexte de visite libre. Ceci nourrit les réflexions sur la conception d'EIAH adaptés au musée pour soutenir l'apprentissage autonome des visiteurs-apprenants.

Mots-clés : expérience des visiteurs, processus d'apprentissage, éducation informelle.

Abstract. Our study about the visit experience is based on stimulated recall interviews with 52 visitors of French and Belgian museums. The access of micro-dynamics of their experience allows to know theirs cognitive and sensory processes used on site. This analysis eventually sheds light on the learning processes that are fostered by the self-tour context. This contributes to the reflections on the design of learning devices adapted to museum that aims to support autonomous learning of learners-visitors.

Keywords: visitor experience, learning processes, informal education.

1 Introduction

La mission d'éducation des musées est inscrite dans la plupart des textes officiels qui définissent leur rôle [1]. Ainsi, selon la définition de référence issue de l'*International Council of Museums* (ICOM), les musées « offrent à leurs publics des expériences variées d'éducation, de divertissement, de réflexion et de partage de

connaissances » [2]. Bien que la dimension éducative des musées semble communément admise, la manière dont elle s'opère reste encore peu documentée. Les recherches de Falk et Dierking constituent la référence dans ce domaine. Ils réalisent plusieurs études de terrain pour évaluer les apprentissages en sortie de visite et proposent des profils de visiteurs déterminés d'après leurs motivations à venir au musée [3].
En parallèle, les technologies sont de plus en plus sollicitées au musée, que ce soit pour dynamiser les visites ou pour pallier l'absence d'un guide. Toutefois, les recommandations pour la conception de dispositifs numériques prennent rarement en compte l'expérience de visite située puisque peu de recherches l'étudient en réalité [4(p. 46)]. L'émergence d'une nouvelle épistémologie de recherche et de moyens techniques modernes permet aujourd'hui d'avoir accès à cette expérience de visite. Ainsi, depuis 2010, Schmitt *et al*. [4] tentent de comprendre en profondeur l'expérience des visiteurs grâce à la méthode REMIND. Riches aujourd'hui de plus de 400 entretiens, leurs recherches donnent accès à des informations de qualité qui permettent de saisir les dynamiques de l'expérience de visite non guidée.
Nous réutilisons aujourd'hui une partie de leur corpus pour interroger la façon dont les visiteurs apprennent via le média exposition. En effet, l'exposition relève d'un procédé de médiatisation, au sens de Peraya [5], d'un discours institutionnel en vue de transmettre des savoirs scientifiques à un public étendu. Le média expositionnel est « spatial, objectal et social » [6], selon Davallon, mais se construit de plus en plus avec des dispositifs numériques qui proposent autant de formes d'environnements pour l'apprentissage. Nos résultats, bien qu'éclairant différents aspects de l'expérience de visite, permettent de faire émerger des pistes utiles pour la conception d'EIAH dans le cadre d'activités d'éducation informelle[1] comme les sorties au musée.

2 Problématique et méthode d'accès à l'expérience

Notre étude porte sur l'analyse de 52 entretiens de visiteurs de 5 musées français et belges : en Belgique, le *Muséum des Sciences naturelles de Bruxelles*, le *Musée du Doudou*, et à Strasbourg, le *Musée de l'Œuvre-Notre-Dame*, *le Vaisseau* et *le Musée Zoologique de Strasbourg*. Ces entretiens ont été réalisés selon la méthode REMIND [4] et sa variante REMIND-Contrast[2], issues des études en *stimulated recall*. Leur principe général repose sur la stimulation de la mémoire du sujet par le visionnage de la vidéo de son activité. Le visiteur est équipé d'*eye-trackers* pour parcourir l'exposition en autonomie. Lors de l'entretien, le participant est amené à décrire et commenter son expérience *in situ* par le rejeu de la vidéo subjective.

[1] Education informelle et non-formelle font l'objet de définitions voisines. La visite libre relève de l'éducation informelle dans la mesure où l'apprentissage n'est pas encadré par un enseignement organisé et intervient dans un parcours autonome dont la dynamique comprend de nombreux échanges interpersonnels. Cf. p.ex. Garnier [7].

[2] REMIND-Contrast combine la méthode REMIND avec le principe de la Repertory Grid (RepGrid) issue de la théorie de Kelly [8]. La RepGrid consiste à faire émerger de l'expérience personnelle des attributs groupés en pôles contraires pour caractériser les situations vécues et décrites par la personne.

Notre étude vise à décrire les types d'engagements des visiteurs avec les dispositifs de médiation pour identifier leurs processus d'apprentissage spontanés en visite libre. Pour répondre à cette problématique, nous avons réalisé une analyse de contenu par « catégories conceptualisantes » [9]. Elle consiste à catégoriser les engagements des visiteurs envers le contenu de médiation, objet potentiel d'apprentissage. Outre les engagements, ces catégories révèlent différents processus cognitifs, sensoriels et sociaux qui renseignent les modes d'apprentissage des visiteurs. Notre approche se concentre dès lors sur la compréhension de la dynamique d'expérience, plutôt que sur l'évaluation d'acquis prédéterminés. En effet, le contexte d'éducation informelle et de l'apprentissage autonome étend les possibilités d'apprentissage bien au-delà de l'intention des concepteurs d'exposition [10]. Ainsi, il nous est apparu avant tout utile de mieux saisir les *processus* d'apprentissage adoptés par les visiteurs en situation de visite libre.

3 Résultats d'analyse et processus d'apprentissage au musée

Les activités récurrentes en visite libre sont ici mises en lien avec les processus d'apprentissage. Nous nous référons aux processus[3] définis par Musial et Tricot [11] en 2020.

L'expérience de visite apparaît avant tout comme une **exploration** de l'espace d'exposition. Cette découverte passe généralement par l'identification de certains expôts d'intérêt pour le visiteur : qu'est-ce qui est exposé, quels sont les attributs de tel ou tel objet ? Il s'agit alors de les **comparer** et de répertorier les particularités de chacun. Ces tâches sont en fait associées au processus de ***conceptualisation (a)*** : former un concept selon ses attributs, l'étiqueter et le mettre en relation avec d'autres. Les informations fournies par les textes de médiation soutiennent largement ce processus en donnant un nom, ou un titre, et en pointant les éléments particuliers de l'expôt[4]. Les visiteurs sont régulièrement à la recherche de ces informations pour identifier mais aussi comparer les expôts entre eux.

Le média exposition apporte par ailleurs des informations de natures diverses, notamment d'ordre sensoriel. La **confrontation physique** avec des objets matériels, la manipulation de reproductions, les visualisations animées ou encore les installations immersives sont autant d'occasions de découvrir autrement des objets ou des phénomènes. Ces dispositifs offrent des points de vue nouveaux et éveillent les sens des visiteurs qui peuvent alors mieux **se représenter** la réalité d'un animal, de l'usage d'un objet, de l'ambiance d'un événement ou encore de l'effet d'un phénomène. Ces représentations mentales, nourries de ces nouvelles informations, permettent en réalité de soutenir leur processus de ***compréhension (b)***.

Les visiteurs relèvent d'ailleurs l'importance des **narrations et remises en contexte** des expôts afin de les aider à se représenter, ou s'immerger dans, des événements absents. Le processus de représentation est, dans le contexte du musée, grandement

[3] Les processus de Musial et Tricot sont numérotés de *(a)* à *(c)* dans le texte.

[4] L'expôt réfère à l'ensemble du matériel expositionnel qui peut être présent au musée comme élément de médiation. Le terme désigne ainsi autant les dispositifs que les œuvres, artefacts ou reproductions afin de s'adapter aux collections de diverses natures [12(p. 223)].

enrichi par les informations sensorielles ou émotionnelles émanant de ces histoires. Le contexte d'exposition apparaît de façon générale comme efficace pour transmettre des données provenant des sens et des émotions qui dépassent le cadre des connaissances formalisées, scolaires ou scientifiques. Ces contextes d'enseignement formel ne prennent généralement que peu en compte les éléments subjectifs ou liés au ressenti.

Enfin, les interactions sociales entre visiteurs encouragent l'engagement avec le matériel expositionnel. Premièrement, leurs échanges servent une forme de **médiation intra-groupe**. Les visiteurs font de nombreuses références face aux expôts que ce soit à la culture populaire (dessins animés, films, légendes, etc.) ou à des expériences personnelles passées (rencontre avec un animal, souvenir de voyage, etc.). La visite est également l'occasion de partager des compétences et des connaissances sur le thème de l'exposition (fait historique, identification d'un objet, d'un artiste, etc.). Ainsi, cette médiation entre visiteurs soutient le processus de ***compréhension (b)***, en apportant des éléments différents de ce que le discours institutionnel communique. Deuxièmement, les visiteurs collaborent régulièrement lors de leur visite. Cette **collaboration** concerne les tâches à réaliser en visite, soit pour l'exploration globale de l'exposition, soit au moment d'interagir avec un dispositif. Les visiteurs se répartissent notamment les tâches liées à la sélection des expôts à examiner et à la lecture des informations. Ils peuvent également collaborer pour résoudre une énigme ou pour comprendre le mode d'emploi d'un dispositif interactif. Ainsi, la visite en groupe est l'opportunité de travailler la ***procéduralisation (c)*** de méthodes de recherche d'information et plus globalement celles d'exploration d'une exposition.

4 Qu'en retirer pour la conception d'EIAH ?

4.1 Structurer, visualiser et synthétiser

Les EIAH pourraient soutenir les processus cognitifs et sensoriels des visiteurs et enrichir ainsi l'expérience de visite. Tout d'abord, les EIAH offrent un espace étendu pour **compléter les informations fournies** par l'exposition. Ainsi, il ne s'agit pas forcément d'ajouter quantité d'informations mais plutôt de la structurer. Les synthèses et visualisations permettraient de regrouper des informations selon un fil conducteur ou une narration. Le visiteur pourrait alors s'épargner un travail fastidieux de récolte des informations réparties sur différents supports de médiation. Ces synthèses donnent par ailleurs la possibilité de croiser et comparer différents expôts plus aisément. Ceci aiderait plus globalement les visiteurs-apprenants à la conceptualisation des connaissances. Par exemple, expliciter les caractéristiques qui font appartenir telle espèce animale à telle famille ou mettre en parallèle des œuvres qui sont représentatives d'une évolution artistique en pointant les éléments qui les différencient.

Les visualisations et synthèses pourraient également s'attacher à indiquer ce qui n'est pas présent dans le musée. La sélection d'expôts est inhérente à la conception d'exposition et mène à ne présenter qu'une partie de ce qui existe. Ceci impacte dès lors les représentations que construisent les visiteurs sur un sujet. Par exemple, le contenu de l'EIAH pourrait rendre compte de la proportion réelle d'individus dans une

population donnée ou de la rareté d'un artefact à une époque. Quel est le nombre de renards blancs dans leur habitat naturel ? Combien de sculptures religieuses ornaient à l'origine la Cathédrale de Strasbourg ? Ainsi, le visiteur pourrait mettre en perspective son impression concernant certains expôts, qu'il rencontre en visite, avec la réalité connue et étayée de ceux-ci.

4.2 Interactivité, interactions et immersion

Mais les EIAH ont l'intérêt plus particulier de pouvoir ajouter une **couche d'interactivité** à la visite. Bien que cette plus-value potentielle soit souvent mise en avant, la question des moyens pour la mettre en œuvre reste un point délicat. L'étude de l'expérience de visite semble permettre de dessiner quelques lignes directrices pour cette conception. Notamment l'apport de la collaboration et de la médiation intra-groupe pour l'apprentissage au musée incite à encourager les interactions sociales en visite libre. D'une part, elles mènent à l'engagement des visiteurs dans l'exposition et, d'autre part, soutiennent les processus de compréhension des visiteurs vis-à-vis du contenu de médiation. Les EIAH peuvent jouer un rôle de choix pour amener les visiteurs à échanger et collaborer de façon structurée et continue tout au long de la visite. L'interaction proposée par l'EIAH orienterait alors leurs échanges pour servir directement leurs apprentissages. Ainsi, la forme ludique apparaît comme une option aux multiples avantages [13] pour la visite au musée. Elle permet par exemple de définir des objectifs communs et de répartir des rôles parmi les visiteurs. L'EIAH pourrait alors structurer les échanges entre visiteurs pour viser des objectifs d'apprentissage plus explicites.

Au-delà de l'interaction entre visiteurs, les EIAH peuvent également **enrichir l'expérience sensorielle** par l'immersion. L'immersion, nous l'avons vu, n'exige pas forcément un environnement virtuel mais peut être suscitée par la mise en narration et la contextualisation. Si le musée donne à voir des objets authentiques, la mise en contexte dans leur environnement d'origine est limitée. Les EIAH pourraient alors participer à la reconstruction de ce contexte et présenter une narration dans laquelle les visiteurs peuvent être acteurs. Cette expérience interactive nourrirait alors les perceptions sensorielles des visiteurs et les aiderait à adopter d'autres points de vue sur un contexte peu ou pas familier. La contextualisation et la mise en rapport avec une histoire humaine servent généralement à offrir de nouvelles perspectives à propos des expôts (usages, valeurs symboliques, émotions rattachées, etc.). Ceci encourage les visiteurs à se représenter une réalité complexe et la représentation constitue en soi un levier intéressant pour le processus de compréhension.

5 Perspectives

L'étude de l'expérience de visite, selon la perspective personnelle des visiteurs, apporte un nouveau regard sur les modalités d'apprentissage au musée. Elle met en exergue une série d'éléments qui concernent la dynamique sociale, les processus cognitifs et les expériences sensorielles favorisés par la visite d'une exposition. Ceci nous permet dès lors de dégager des lignes directrices intéressantes pour la conception d'EIAH adaptés

au contexte de visite libre. Il apparaît que les EIAH peuvent soutenir les processus d'apprentissage en visite de plusieurs manières. Nous en avons pointé deux principales : la structuration de l'information et la nature de l'expérience à vivre au musée. Notre étude questionne plus largement l'usage des technologies dans les musées et leurs intérêts particuliers dans ce contexte. Les retours d'expériences des visiteurs mettent en effet l'accent sur l'apport de la sensorialité et de la sociabilité pour s'engager avec le contenu de médiation. Notre analyse montre que ces éléments peuvent directement servir l'expérience d'apprentissage autonome et marquent les visiteurs de façon positive. Alors si les EIAH peuvent structurer le contenu de médiation, ils peuvent également améliorer les composantes sensorielles et sociales du contexte de visite pour en enrichir l'expérience d'apprentissage.

References

1. Rivet, M.: La définition du musée : Que nous disent les droits nationaux ? In : Mairesse, F. (dir.), Colloque juin 2017, Définir le musée du XXIe siècle, pp. 53-123. ICOFOM, Paris (2017).
2. ICOM « Définition du musée », https://icom.museum/fr/ressources/normes-et-lignes-directrices/definition-du-musee/, dernière consultation 10/05/2023.
3. Falk, J.: An Identity-Centered Approach to Understanding Museum Learning. Curator : The Museum Journal, 49, 151-166 (2010).
4. Schmitt, D., Aubert, O.: REMIND : A method to understand the micro-dynamics of the museum visitors experience. RIHM, 17(2), 43-70 (2017).
5. Peraya, D.: Un regard critique sur les concepts de médiatisation et médiation : Nouvelles pratiques, nouvelle modélisation. Les Enjeux de l'information et de la communication (2008).
6. Davallon, J.: Pourquoi considérer l'exposition comme un média ? Médiamorphoses, 9, 27-30 (2003).
7. Garnier, B.: L'éducation informelle contre la forme scolaire ? Carrefours de l'éducation, 45(1), 67-91 (2018).
8. Kelly, G. A.: The Psychology of personal constructs, vol. 1. © 1955. Norton &Co, New-York (1991).
9. Paillé, P., Mucchielli, A.: L'analyse qualitative en sciences humaines et sociales. 5e éd. Armand Colin, Malakoff (2021).
10. Schmitt, D.: Ce que "comprendre" signifie pour les jeunes visiteurs dans un centre de culture scientifique. Questions de communication. Série actes, 25, 225-236 (2015).
11. Musial, M., Tricot, A.: Précis d'ingénierie pédagogique. De Boeck Supérieur, Louvain-la-Neuve (2020).
12. Desvallées, A.: Cent quarante termes muséologiques ou petit glossaire de l'exposition. In : de Bary, M.-O., Tobelem J.-M., Manuel de muséographie, pp. 205-251. Séguier, Biarritz (1998).
13. Sauvé, L., Renaud, L., Gauvin, M.: Une analyse des écrits sur les impacts du jeu sur l'apprentissage. Revue des sciences de l'éducation, 33(1), 89-107 (2007).

Vers la compréhension des difficultés de lecture en L2 à travers des paramètres acoustiques et de mouvement des yeux

Sofiya Kobylyanskaya[1,2], Ioana Vasilescu[1,2], Laurence Devillers[1,2,3] et Olivier Augereau[4]

[1] LISN-CNRS, Orsay, France
[2] Université Paris-Saclay, Orsay, France
[3] Sorbonne-Université, Paris, France
[4] Lab-STICC, ENIB, Brest, France

Résumé. L'étude porte sur l'analyse de l'acquisition de l'anglais langue étrangère L2 à travers deux modalités : la parole et le mouvement des yeux lors d'une expérience de lecture à voix haute. Les expériences ont été menées en France et au Japon ce qui permet de mettre en évidences les influences de la L1 et de la culture sur les performances en L2. Nous présentons les premières tendances observées sur les deux axes d'analyse et proposons des directions pour de futures recherches.

Mots-clés : L2 , multimodalité , analyse acoustique , mouvement des yeux.

1 Introduction

LeCycl (Learning Cyclotron) est un projet ANR trilatéral entre la France, le Japon et l'Allemagne dont le but est de renforcer le processus d'acquisition et de transfert des connaissances, et plus particulièrement de l'anglais langue étrangère (L2). L'objectif de l'étude (au sens large) est de trouver des corrélations entre les performances en anglais L2, des habitudes de l'enseignement spécifiques à la culture du pays de l'apprenant (la France vs le Japon) et des mesures objectives provenant de l'analyse acoustique et du mouvement des yeux. Afin d'atteindre cet objectif, dans un premier temps, nous nous focalisons sur l'analyse des zones textuelles difficiles lors de la lecture à voix haute. Nous considérons qu'une zone difficile est un extrait de texte d'au moins un mot pouvant contenir un mot inconnu marqué par les participants et/ou dont la lecture se manifeste par des disfluences (pauses, hésitations, reprises). Ensuite, en déterminant les difficultés de lecture, nous envisageons d'implémenter des nudges [13] adaptés aux particularités individuelles des apprenants en leur suggérant des stratégies d'amélioration de leur technique de lecture. Dans cet objectif, trois protocoles expérimentaux (P1,P2,P3) [Table 1] de lecture à voix haute ont été mis en place en France[9] (P1) et au Japon[12] (P2, P3). Dans cet article, nous nous focalisons davantage sur les parties réalisées par l'équipe française (P1, P3 partie 1), comparons les protocoles, décrivons les mesures utilisées et les étapes d'analyse.

Table 1. Protocoles expérimentaux

	Protocol 1	**Protocol 2**	**Protocol 3 (joint)**
Lieu	France	Japon	Japon
Équipe	LISN-CNRS	OMU	LISN-CNRS+OMU
# participants	50	20	20
L1	français	japonais	japonais
Tâche	3 txt, QCM, mots inconnus	mangas, mots inc.	P1+P2+manga sans img
Équipement	micro, eye-tracker	micro, eye-tracker	micro, eye-tr.
Durée	30 min	3h	1h : txt+manga sans img 2h : mangas avec img

2 Contexte et hypothèses

De nombreuses études abordent la question de l'acquisition d'une L2 via la modélisation de la parole ou des mesures du mouvement des yeux. Nous proposons de fusionner les deux méthodes, l'apport consistant en des expérimentations qui s'appuient sur les deux modalités tout en prenant en compte le facteur culturel. D'un point de vue linguistique, la réalisation des phonèmes de la L2 dépend de celles de la L1 et s'explique par le phénomène du "crible phonologique" [4]. D'un point de vue du mouvement des yeux, on peut estimer la compréhension du texte lu en L2 [1] mais on obtient peu d'informations sur les indices effectifs. Plusieurs études montrent la complémentarité de ces modalités, par exemple, en mettant en avant la corrélation entre la planification de la parole et le mouvement des yeux ainsi qu'entre le temps de l'acquisition de la forme phonologique du mot et sa longueur [7].

Le but de notre étude est d'estimer les difficultés de la lecture à voix haute des textes en L2 en nous basant sur ces métriques et de trouver des corrélations entre les résultats vs L1, l'environnement de l'apprentissage [10], les habitudes de l'enseignement spécifiques à la culture du pays de l'apprenant [5], le type de support d'enseignement [11]. Concrètement, nous vérifions si les différences du système linguistique du français et du japonais font émerger des passages textuels difficiles différents pour les représentants des deux cultures : par exemple, des mots contenant des "th" ou des "h" pour les francophones et les mots contenant des clusters consonantiques pour les japonophones. En plus, nous souhaitons étudier la stratégie dont les participants abordent les difficultés de prononciation ou de compréhension en fonction de leur système éducatif : la lecture syllabique, la recherche du sens dans le contexte gauche/droite/en haut/en bas, des pauses silencieuses/remplies (hésitations), des répétitions, le changement du degré de l'expressivité de la lecture etc. Notre étude porte également sur l'impact du support d'apprentissage sur les performances de lecture et de compréhension. Par exemple, l'utilisation des images et des mangas dans l'enseignement a été efficace dans plusieurs cas [11,14]. Nous explorons plus particulièrement l'impact

du texte vs texte+images (manga), le dernier étant propre à la culture japonaise y compris en situation d'apprentissage.

3 Méthodologie

Public. L'expérience a été menée auprès d'un public adulte francophone et japonophone natif fait majoritairement d'étudiants âgés de 18 à 30 ans (85% en France et 100% au Japon) et ayant le niveau A1-C2 en anglais. Au moment de la rédaction de l'article, nous avons recueilli des données auprès de 50 participants en France (P1) et 20 au Japon (P3).

Équipement. Pour l'enregistrement des participants, nous utilisons le microphone AKG Perception Wireless 45 Sports Set Band-A 500-865 MHz, l'eye-tracker Tobii Nano Pro et le système d'enregistrement du mouvement des yeux et de la voix Eye Got It [2], comme dans [9].

Support Deux types de support sont mis en place : des textes en anglais de différents niveaux [9] (P1) et des mangas en anglais traduits du japonais [12] (P2). Le P3 est une fusion des P1 et P2 avec l'utilisation de deux supports comme dans P1 et P2 et d'un extrait de manga sous format texte. Cela permet de comparer les stratégies de lecture et le niveau de compréhension de textes accompagnés ou non d'images. Le choix d'un extrait de manga pour P3 a été basé sur des critères linguistiques. Du point de vue syntaxique, les mangas étudiés contiennent majoritairement des phrases courtes, sans subordonnées relatives ni temps verbaux complexes, car ils représentent le discours direct interactif qui se caractérise par des structures grammaticales simples. Ainsi, les critères suivants ont été pris en compte : la présence du lexique du niveau avancé (C1) permettant d'avoir des supports de niveaux comparables ; la présence du champ lexical (en l'occurrence de la nourriture) afin de faire référence au texte du niveau avancé du P1 ; la présence des clusters consonantiques afin de créer des conditions phonologiques complexes pour les participants japonophones dont la langue maternelle ne comprend que des syllabes type consonne-voyelle (CV).

4 Méthode d'analyse

L'objectif courant de l'étude est d'identifier des zones textuelles qui soulèvent des défis linguistiques de sources diverses : phonologique, lexicales, syntaxiques et de mettre en évidence des caractéristiques acoustiques ou physiologiques de leur réalisation par rapport aux zones non marquées comme difficiles (par exemple, le débit ralenti, le nombre de saccades/fixations augmenté/diminué). Nous présentons ensuite les démarches d'analyse des deux modalités : la parole et le mouvement des yeux qui sont étudiées séparément à cette étape de travail.

Parole. Tout d'abord, le texte a été aligné avec le son à l'aide du logiciel WebMaus [8]. Le texte a été manuellement enrichi en disfluences (hésitations, répétitions etc.) afin de correspondre à la production du locuteur par rapport au

texte original et de permettre un alignement fidèle à la prononciation. Une fois le texte aligné, des fenêtres d'analyse de 5 secondes ont été extraites avec les intervalles d'une seconde. La durée de la fenêtre est suffisante pour l'extraction des 88 caractéristiques acoustiques avec OpenSMILE eGeMAPS [3] et 14 caractéristiques prosodiques myprosody [5] et d'observer leur variation d'un segment à l'autre. Les mêmes caractéristiques ont été extraites sur le texte entier afin de trouver des valeurs moyennes qui nous servent de référence. Une autre référence est la lecture d'un locuteur natif de l'anglais américain.

Mouvement des yeux. Les caractéristiques du mouvement des yeux ont été extraites également au niveau de la lecture du texte entier et au niveau des extraits de 5 secondes ce qui permet de faire la corrélation avec les caractéristiques acoustiques. De la même manière, les valeurs moyennes du participant et d'un locuteur natif servent de référence. Les valeurs suivantes ont été extraites : le nombre de saccades/fixations, la durée minimale/moyenne/maximale des saccades/fixations, la déviation standard des durées des saccades/fixations.

5 Résultats préliminaires

Cette section est dédiée aux analyses des premières tendances qui émergent pour ce qui est des deux modalités, parole et mouvement des yeux. Pour l'instant nous avons mis en oeuvre ces mesures sur les données collectées en France (50 participants) et nous évaluerons par la suite les différences culturelles. Nous considérons des valeurs moyennes d'un ensemble de valeurs de mouvement des yeux et de parole en vue de corrélations entre les caractéristiques extraites, le niveau L2 des participants et le niveau des textes lus. En nous basant sur ces observations, nous pourrons ensuite nous focaliser sur la recherche des extraits de textes difficiles et les caractéristiques de leur réalisation.

Mouvement des yeux. Au niveau du mouvement des yeux, on remarque que le nombre et la durée totale des fixations diminue avec l'augmentation du niveau de la L2 [Fig.1]. Les participants des niveaux B1 et B2 effectuent plus de saccades et elles sont plus longues que celles des participants des niveaux C1 et C2. Par contre, les saccades sont moins nombreuses et plus courtes pour les participants du niveau A2. Cette tendance nécessite d'être vérifiée auprès d'un plus grand nombre de participants, car seulement 3 participants de notre corpus ont le niveau A2. Les participants des niveaux B1 et B2 utilisent davantage de saccades de régression en lisant les textes des niveaux intermédiaire et avancé. L'hypothèse peut être formulée que les débutants, confrontés à des problèmes de compréhension, se focalisent plutôt sur la prononciation que sur la compréhension, alors que les participants plus avancés comprennent le texte et ont besoin de faire moins de retours en arrière. Concernant les participants intermédiaires, nous supposons que leur niveau leur permet d'avoir une bonne compréhension des textes, mais nécessite des retours en arrière occasionnels afin de clarifier le contexte ou la structure syntaxique.

5. https ://github.com/Shahabks/myprosody

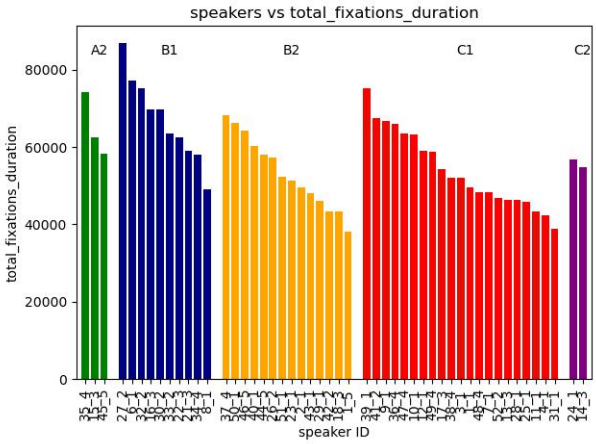


Fig. 1. Durée totale de fixations par participant et par niveau pour le texte avancé. Axe x - l'ID du participant, Axe y - la durée des fixations

Parole. Du point de vue de la parole, une attention particulière est portée aux variations de la fréquence fondamentale (F0), au débit de la parole et d'articulation, la durée de la parole, le nombre des pauses et de syllabes, car ces caractéristiques sont parmi d'autres qui sont distinctives de la fluence en L2 [6]. On observe une baisse de la F0 avec la progression de la L2 qui peut être due à des raisons linguistiques telles que l'impact de la complexité du texte ou extra-linguistiques, telles que les émotions (la montée du stress), le genre, les habitudes culturelles. Dans la plupart des cas (sauf pour A2), le nombre de pauses et de syllabes diminue avec l'avancement du niveau L2. On suppose que les participants ayant des niveaux de langue plus élevés font moins de disfluences lors de la lecture (moins de répétitions et d'hésitations) et donc, moins de syllabes sont détectées par le système automatique. L'intensité de la voix et le nombre de segments voisés (où la F0 est non nulle) augmentent avec le niveau L2 sauf pour le C2 pour lequel nous n'avons que 2 participants. Cela peut refléter la montée du niveau de confiance avec l'avancement de la L2, mais aussi des particularités personnelles non liées à la maîtrise de la langue étrangère.

6 Conclusions et discussion

L'article présente les résultats préliminaires d'une analyse multimodale de l'acquisition de la L2 et propose des pistes pour de futures analyses des difficultés de lecture liées à des particularités linguistiques, culturelles et personnelles. Les premières tendances observées confirment l'intérêt d'une analyse jointe des paramètres acoustiques et du mouvement des yeux pour l'étude de l'apprentissage de la L2. En effet, nous observons des corrélations des paramètres étudiés avec le niveau L2 des participants et le niveau de difficulté des textes lus. Ces ten-

dances nécessitent une validation statistique plus approfondie tout comme des données supplémentaires pour confirmer les tendances. A cet effet, nous avons prévu d'augmenter le nombre de participants ayant les niveaux A2 et C2 pour équilibrer la base de données. Les futures étapes du travail seront dédiées à l'extraction des mêmes caractéristiques sur le corpus collecté au Japon et la recherche des tendances, au calcul des paramètres distinctifs des zones textuelles selon un critère de difficulté permettant de corréler et de prédire les contextes problématiques et d'adopter des stratégies de remédiation adaptées.

Références

1. Augereau, O., Fujiyoshi, H., Kise, K. : Towards an automated estimation of english skill via toeic score based on reading analysis. pp. 1285–1290 (12 2016)
2. El Baha, M., Augereau, O., Kobylyanskaya, S., Vasilescu, I., Devillers, L. : Eye got it : a system for automatic calculation of the eye-voice span (2022)
3. Eyben, F., Wöllmer, M., Schuller, B. : Opensmile : The munich versatile and fast open-source audio feature extractor. Association for Computing Machinery, New York, NY, USA (2010)
4. Flege, J. : Second language speech learning : Theory, findings and problems, pp. 229–273 (01 1995)
5. Grainger, P. : The impact of cultural background on the choice of language learning strategies in the jfl context. System **40**(4), 483–493 (2012)
6. Hsieh, C.N., Zechner, K., Xi, X. : Features Measuring Fluency and Pronunciation, chap. 6, pp. 101–122 (11 2019)
7. Huettig, F., Rommers, J., Meyer, A. : Using the visual world paradigm to study language processing : A review and critical evaluation. Acta psychologica **137**, 151–71 (06 2011). https://doi.org/10.1016/j.actpsy.2010.11.003
8. Kisler, T., Reichel, U., Schiel, F. : Multilingual processing of speech via web services. Computer Speech & Language **45**, 326–347 (2017)
9. Kobylyanskaya, S. : Speech and eye tracking features for l2 acquisition : A multimodal experiment. In : Rodrigo, M.M., Matsuda, N., Cristea, A.I., Dimitrova, V. (eds.) Artificial Intelligence in Education. Posters and Late Breaking Results, Workshops and Tutorials, Industry and Innovation Tracks, Practitioners' and Doctoral Consortium. pp. 47–52. Springer International Publishing, Cham (2022)
10. Nikitina, L. : Creating an authentic learning environment in the foreign language classroom. International Journal of Instruction **4**(1) (2011)
11. Ranker, J. : Using comic books as read-alouds : Insights on reading instruction from an english as a second language classroom. The Reading Teacher International Reading Association **61**, 296–305 (12 2007)
12. Takaike, T., Iwata, M., Kise, K. : Estimation of unknown words using speech and eye gaze when reading aloud comics. In : MANPU 2022 The 5th International workshop on comics analysis, processing and understanding (2022)
13. Thaler, R., Sunstein, C. : Nudge : Improving decisions about health, wealth, and happiness. Yale University Press (2008)
14. Themelis, C., Sime, J.A. : Comics for inclusive, technology-enhanced language learning (2020)

EIAH2023 : 11ème Conférence sur les Environnements Informatiques pour l'Apprentissage Humain
La conférence pluridisciplinaire francophone sur la conception et l'analyse des environnements numériques pour l'éducation et la formation
12-16 juin 2023 Brest (France)

Session de communications 1.2
Conception et outils pour la conception

Outiller la conception participative et générative de tableaux de bord d'apprentissage

Madjid Sadallah et Jean-Marie Gilliot

IMT Atlantique, Lab-STICC UMR CNRS 6285, F-29238 Brest, France
madjid.sadallah@imt-atlantique.fr, jm.gilliot@imt-atlantique.fr

Résumé. Les tableaux de bord d'apprentissage visent à soutenir la prise de décision au cours du processus d'apprentissage. Cependant, leur adoption à grande échelle reste limitée. Si ce constat peut s'expliquer par leur apparition relativement récente, la recherche montre que ceci est également dû au manque de participation des parties prenantes au processus de conception. Afin de soutenir ce processus tout en impliquant les utilisateurs, nous proposons d'articuler autour d'un espace de conception partagé un outil de conception participative pour la phase d'idéation et un outil génératif pour la phase de prototypage. Nous analysons ces outils en termes de facilité d'utilisation, de soutien à la participation ou d'expérience utilisateur, suivant les phases. L'analyse des retours obtenus montrent que ces outils apportent un soutien tant à la participation des utilisateurs finaux qu'à l'expérience utilisateur des développeurs.

Mots-clé : Tableaux de bord d'apprentissage · Conception participative · Conception générative · Génération de tableaux de bord.

Abstract. Learning dashboards are intended to support decision making during the learning process. Yet, their widespread adoption remains limited. While this can be explained by their relatively recent appearance, research shows that this is also due to the lack of stakeholder participation in the design process. In order to support this process while involving users, we propose to articulate around a shared design space a participatory design tool for the ideation phase and a generative tool for the prototyping phase. We analyze these tools in terms of usability, participation support or user experience, depending on the phases. The analysis of the feedback obtained shows that these tools support both the participation of the end-users and the user experience of the developers.

Keywords: Learning Dashboards · Participatory design · Generative design · Dashboard generation.

1 Introduction

Les tableaux de bord d'analyse de l'apprentissage (TBA) ou *Learning Analytics Dashboards* sont des dispositifs qui organisent et présentent via des visualisations de données des indicateurs liés aux apprenants, aux processus d'apprentissage

et/ou aux contextes d'apprentissage [20]. Leur objectif est de communiquer des informations provenant des données des apprenants aux acteurs éducatifs afin que ces derniers puissent prendre des décisions et de mener des interventions judicieuses ayant un impact significatif sur le processus d'apprentissage [11]. La littérature actuelle, tout en montrant un intérêt croissant pour les TBA (voir par exemple [20]), relève le fait que leur diffusion à grande échelle est encore limitée [1], en raison notamment de la rareté des études sur leurs principes de conception [7] et de la faible implication des parties prenantes dans le processus de conception [2]. Dans ce contexte, la *conception participative* (ou *codesign*), dérivée de la conception centrée sur l'utilisateur, fait l'objet d'un intérêt croissant dans la communauté [14]. Une telle approche favorise la construction d'un consensus, de modèles mentaux partagés et d'utilisations appropriées des informations par les différentes parties prenantes. Bien que la pertinence de cette approche de conception soit reconnue et que quelques exemples de son utilisation en éducation soient rapportés dans la littérature [18], son utilisation reste relativement rare [18], notamment parce que les approches pour y parvenir sont encore floues [2] et que la communauté ne dispose pas d'outils adaptés à ses besoins permettant de la mettre en œuvre de manière efficace [2, 6]. Partant de ce constat, et dans le cadre d'un projet qui vise à élaborer et à outiller un processus de conception de TBA, nous proposons une suite composée d'un outil de conception participative pour la phase d'idéation et un outil génératif pour la phase de prototypage, autour d'un espace de conception partagé. Pour valider l'adéquation de cette suite, nous analysons ces outils en termes d'utilisabilité, de soutien à la participation pour la phase d'idéation et d'expérience utilisateur pour la phase de prototypage.

Cette recherche a été menée suivant une méthodologie *Design-Based Research* (DBR) [23], une approche itérative largement utilisée pour examiner les environnements éducatifs à travers des cycles de conception, de mise en œuvre et d'analyse. La première itération [8] a permis de valider la démarche de conception et la faisabilité d'articuler un outil d'idéation et un outil de prototypage autour d'un espace de conception [21]. L'itération actuelle vise à (1) *raffiner l'espace de conception* en réifiant le processus de prise de décision ; (2) proposer un *outil d'idéation PaDLAD* permettant d'exprimer le processus de prise de décision et facilitant la coopération ; (3) développer un *outil de prototypage opérationnel LADStudio* ; et (4) analyser les outils développés. La prochaine section présente la méthodologie de conception et le modèle de TBA qui la supporte. Les outils conçus pour outiller la conception participative du TBA font l'objet d'une seconde section, suivie par une section qui décrit les expérimentations réalisées pour analyser ces outils et valider leur utilisation.

2 Conception participative et générative de TBA

Méthodologie de conception. Nous nous basons sur une méthodologie qui combine deux approches complémentaires : *conception participative* et *conception générative*. La conception participative en éducation est une approche où les apprenants, les éducateurs, les institutions, les chercheurs et les développeurs

sont associés aux différentes étapes du processus de conception, de l'exploration à la mise en œuvre réelle [14]. La conception générative est un processus dans lequel l'humain dispose d'outils pour décrire ses besoins et son intention, explorer l'espace de conception, générer un ensemble de solutions cibles, puis sélectionner et affiner la plus appropriée en fonction de son propre jugement [12].

Nous nous situons dans le cadre du *modèle de codesign en analytique de l'apprentissage* proposé par [14]. Selon ce modèle, le processus de conception participative peut être décliné en un ensemble d'activités qui sont itérées de manière à affiner les exigences et à se rapprocher de la solution souhaitée. Notre objectif est d'outiller ce processus plus spécifiquement pour les TBA.

Dans toute approche de conception, il est important d'expliciter les rôles des acteurs impliqués. Dans le cadre des TBA, nous pouvons en distinguer plusieurs que nous regroupons sous deux rôles clés : (1) le rôle d'*utilisateur* (enseignants, apprenants, administrateurs, etc) fait référence aux acteurs dont émane l'expression du besoin et qui seront des utilisateurs finaux; et (2) le rôle de *développeur* qui désigne les acteurs en charge de la spécification, la définition des données et des visualisations, la conception de l'interface utilisateur et l'implémentation du TBA.

Espace de conception. La notion d'*espace de conception* [21] identifie les différentes alternatives et les décisions structurantes de la conception, offrant un support pertinent pour l'idéation, la création et l'évaluation. Dans le cas de la conception de TBA, les questions des 5 W's peuvent être utilisées pour aborder un tel espace de conception [5]. Nous formulons les 5 W's comme suit: *Qui ?* indique l'audience et la circulation entre les différents utilisateurs ; *Quand ?* permet de répondre à la question de savoir si l'utilisation est en temps réel ou différé ; *Pourquoi ?* traduit l'objectif du TBA en termes de décision ; *Quoi ?* détaille le contexte d'utilisation du TBA, et les données pertinentes ; et *Comment ?* se rapporte à la visualisation ainsi qu'aux interactions. Le tableau 1 résume les propriétés que nous avons identifiées pour chacune de ces dimensions de conception, et que nous avons détaillées dans [9].

3 Outillage de la conception participative

Pour outiller le processus de conception participative en se basant sur l'espace de conception proposé, nous avons développé deux outils complémentaires : un outil d'idéation collaborative et un outil de prototypage rapide. Les utilisateurs sont la cible de l'outil d'idéation, tandis que pour l'outil de prototypage ce sont les développeurs, collaborant avec les utilisateurs, qui forment la cible.

3.1 Outillage de la phase d'idéation

La phase d'idéation est cruciale car elle implique la participation de parties prenantes avec des perspectives et des connaissances variées pour définir les objectifs de la conception dès le début du processus. Pour soutenir cette phase,

Table 1: Dimensions de l'espace de conception des TBA

Dimension	**Éléments**	**Valeurs**
Qui ?	Utilisateur	Gouvernance, Institution, Cursus, Enseignant/tuteur, Apprenant
	Circulation	Public, organisationnel, social, individuel
Quand ?	Temps réel	O/N
Pourquoi ?	Focus	Processus d'apprentissage ((méta-)cognitif, orienté résultats ou processus, comportemental, social)
		Gestion (personnes, ressources, activités, expérience)
	Niveau visé de conscience de la situation	Perception (ou suivi), Compréhension (ou analyse), Action (ou projection, décision, intervention)
Quoi?	Données	Liste des données pertinentes
	Portée des données	Apprenant, enseignant, classe, institution
	Source de données	Salle de classe, Systèmes de gestion de l'apprentissage, Programme d'études, Profil, Autre
	Durée des données	Une session, Un semestre, Une année, Toute la scolarité, Toute la vie
Comment?	Visualisation	Type de diagramme
	Interaction	Zoom, Filtre, Détails à la demande, Relation, Historique, Extraction

nous avons développé PaDLAD [16], une boîte à outils d'idéation qui utilise des personas, des cartes d'exploration, des panneaux et des supports de sketching pour encourager la créativité et expliciter les différentes dimensions de l'espace de conception. Le processus de conception à l'aide de PaDLAD comprend trois phases représentées par des panneaux dédiés (Figure 1).

Le *Panneau d'identification* permet d'accueillir un *formulaire de persona*, un *formulaire d'objectif* et un ensemble de *cartes de contexte*. Le *formulaire de persona* peut être utilisé pour personnifier les parties prenantes et recueillir leurs informations. La *fiche objectif* permet d'établir le problème que le tableau de bord vise à résoudre. Cet objectif est défini en fonction du niveau de conscience de la situation recherché (suivi, analyse, action). Les *cartes de contexte* décrivent l'utilisation attendue du tableau de bord : les *Cartes de public* permettent de définir le champ d'analyse ; les *Cartes de données* définissent la source de données ciblée et le temps d'observation.

Le *Panneau DataViz (Données & visualisation)* répond à la question *Quoi?* et comporte un espace réservé aux tuples construits à partir de cartes de données et de visualisation. Les cartes *Données* (ou *Mesure*) identifient les données et les indicateurs pertinents pour atteindre l'objectif du tableau de bord. Les cartes *Visualisation* sont un ensemble de cartes technologiques proposant des visu-

(a) Identification (b) DataViz (c) Sketching

Fig. 1: Panneaux PaDLAD

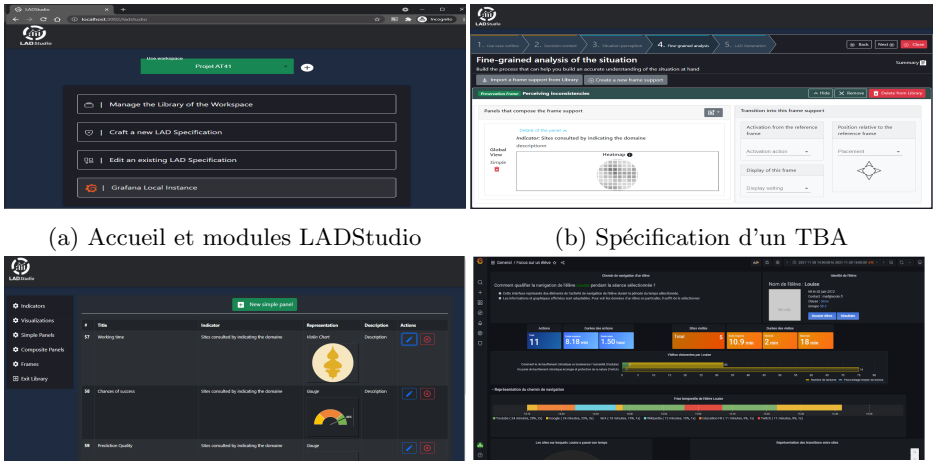
alisations classiques, et qui sont pertinentes pour représenter les informations contenues dans les cartes de données.

Le *panneau Sketching* a pour objectif de permettre aux acteurs de la conception de créer des vues et des représentations graphiques en réalisant des croquis du TBA visé, et de définir des options d'interaction. Nous distinguons trois types de vues : (1) les vues *Perception* permettent de surveiller l'état de l'environnement ; (2) les vues *Compréhension* comportent des représentations visant à fournir l'éclairage nécessaire pour analyser et comprendre une situation donnée ; et (3) les vues *Projection* permettent de préparer l'utilisateur à agir sur les situations découvertes et analysées dans les niveaux précédents.

3.2 Outillage de la phase de prototypage et de support

Pour outiller cette phase, nous avons développé LADStudio [15], un outil qui permet de générer des prototypes de TBA via une spécification déclarative résultant de la phase d'idéation. Il est proposé sous forme d'un service Web intégrant pour le rendu une instance de l'outil de visualisation Grafana.

Le module de spécification permet de décrire un TBA progressivement (Figure 2b). Cette description peut être une traduction des spécifications réalisées durant l'étape d'idéation (par exemple avec PaDLAD). Des écrans séquentiels composent un scénario : (1) description du contexte cible et de l'utilisation visée ; (2) définition de l'objectif principal du TBA ; (3) description des vues de suivi, de compréhension et de projection ; et (4) génération d'une spécification du TBA pouvant être affichée sur le module Grafana (Figure 2d). Une telle spécification peut être rééditée, permettant ainsi un processus cyclique d'édition et de test. Les composants créés durant une spécification sont associés à des interactions qui soutiennent le processus de création de sens. Ils sont automatiquement sauvegardés dans une bibliothèque, ce qui simplifie l'alimentation de celle-ci, permettant la capitalisation et la réutilisation. La bibliothèque peut recevoir différents types de composants : indicateurs, visualisations, panneaux (Figure 2c) et vues. Elle est extensible en permettant de définir de nouveaux composants, de modifier les composants existants et d'en supprimer (en cas de redondance par exemple).



(a) Accueil et modules LADStudio (b) Spécification d'un TBA

(c) Librairie de panneaux (d) Prototype d'un TBA généré

Fig. 2: Quelques éléments de LADStudio

4 Expérimentations

4.1 Problématique

Nous nous intéressons à évaluer les outils proposés comme support à la conception participative de tableaux de bord d'apprentissage. Nos questions de recherche sont les suivantes : *QR1* : Est-ce que les outils proposés sont utilisables par les parties prenantes aussi bien pour la phase d'idéation que de spécification et génération de TBA ? ; *QR2* : L'outil de conception participative encourage-il la collaboration et soutient-il la créativité participative au sein d'un groupe ? ; et *QR3* : l'outil de prototypage assure-il une bonne expérience utilisateur ?

4.2 Méthodologie

Deux études qualitatives ont été menées pour évaluer les deux outils. Quinze participants ont pris part à l'étude de PaDLAD, outil destiné à la phase d'idéation et testé avec des enseignants, chercheurs et concepteurs pédagogiques sans connaissance particulière en EIAH. Treize participants ont participé à l'étude de LADStudio, outil destiné à la phase de prototypage et testé avec des enseignants, chercheurs et concepteurs pédagogiques ayant des connaissances techniques sur les tableaux de bord et le développement d'interface (IHM) lié aux EIAH. Bien que LADStudio fût destiné aux développeurs, des participants ayant des connaissances techniques ont été recrutés en raison de la difficulté de trouver suffisamment de développeurs de TBA pour participer à l'étude. Les données sociodémographiques des participants sont présentées dans le tableau 2.

Table 2: Données démographiques des participants aux deux évaluations

Variable/Catégorie		PaDLAD (N = 15)	LADStudio (N = 13)
Genre	Masculin	07	08
	Féminin	08	05
Âge	30–45	11	07
	46–60	04	03
Profession	Enseignant	06	04
	Concepteur pédagogique	07	03
	Chercheur	02	06

Procédure. Pour évaluer PaDLAD, nous avons organisé un atelier de conception participative de TBA. Après présentation de l'outil, les participants ont été répartis en groupes pour concevoir des croquis de TBA. Cette session a duré environ une heure et demie. L'évaluation de LADStudio, réalisée deux mois plus tard, a débuté par une session de démonstration de l'outil. Les participants ont ensuite été invités à expérimenter l'outil de manière individuelle, puis à participer à un atelier de conception de TBA. L'étude a duré environ deux heures. À la fin de chaque étude, les participants ont été invités à remplir individuellement un questionnaire et à répondre à des questions ouvertes.

Instruments. Nous avons évalué l'utilisabilité des outils PaDLAD et LADStudio moyennant des questionnaires SUS (*System Usability Scale*) [4]. Le SUS est un instrument valide permettant d'obtenir une vision globale de l'utilisabilité des systèmes, et applicable même sur un petit échantillon [22]. En se basant sur le modèle SUS général [4], nous avons préparé pour chaque étude un questionnaire à dix affirmations, chacune mesurée sur une échelle de Likert à 5 points (allant de *Tout à fait en désaccord* à *Tout à fait en accord*).

Pour mesurer le sentiment d'implication perçu par les participants lors des ateliers d'idéation, nous avons utilisé le questionnaire *Self-Report Level of Participation Survey* (SRLPS) proposé dans [10]. Cet instrument est axé sur cinq activités de codesign : (1) Planification et organisation, (2) Processus de conception créative, (3) Définition des priorités, (4) Négociation, et (5) Réflexion et évaluation. Chaque participant doit évaluer son niveau de participation pour chaque activité en utilisant l'échelle suivante : 1 = *passif* (participation de faible niveau), 2 = *partage d'informations*, 3 = *engagement et mobilisation*, 4 = *collaboration*, et 5 = *autonomisation* (participation de haut niveau). L'objectif était d'atteindre pour chaque activité une médiane de 4 (collaboration).

L'évaluation de l'expérience utilisateur (UX) de LADStudio a été réalisée moyennant un questionnaire UEQ (*User Experience Questionnaire*), un outil valide permettant de mesurer de manière exhaustive l'UX de produits interactifs [13], applicable à de petits groupes [19]. Le questionnaire regroupe 26 items répartis en six échelles [17]. Ces échelles évaluent les impressions générales de l'outil (*attraction*), sa facilité d'utilisation et son efficacité (échelles *efficacité* et

compréhensibilité), le sentiment de contrôle de l'utilisateur (*contrôlabilité*), son aspect excitant et motivant (*stimulation*), ainsi que l'originalité et la créativité de la conception de l'outil (*originalité*).

4.3 Résultats

Utilisabilité des outils. Pour analyser les résultats, nous avons calculé les scores SUS normalisés (valeurs entre 0 et 100). Les résultats de l'évaluation des participants des deux outils sont présentés sur la figure 3. La valeur moyenne du score SUS de PaDLAD était de $73,5$ (écart-type de $9,34$) et celle de LADStudio 71.15 (écart-type de 6.15). Parce qu'un score SUS est considéré acceptable s'il est supérieur à 68 [4], nous pouvons affirmer que les deux outils obtiennent un niveau satisfaisant et acceptable d'utilisabilité.

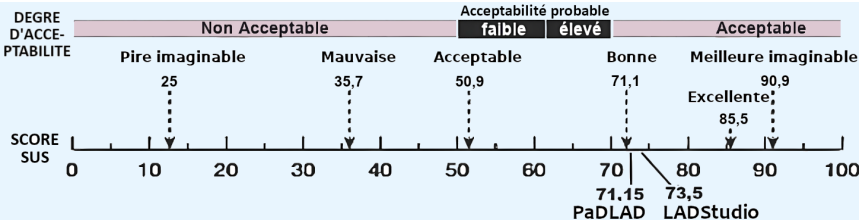


Fig. 3: Résultats de l'étude d'utilisabilité de PaDLAD et de LADStudio.

Les scores SUS peuvent être associés à une échelle d'adjectifs avec degré d'acceptabilité, proposé dans [3], afin de leur attacher une signification plus descriptive. En utilisant cette échelle, comme le montre la figure 3, les scores SUS attribué à ces outils signifient que leur *degré d'acceptabilité* est de niveau *Acceptable*, et que leur niveau sur l'échelle des grades est de catégorie *Bonne*.

Qualité de participation de PaDLAD. Les valeurs obtenues à base du questionnaire SRLPS pour les activités mesurées sont affichées sur le tableau 3. La note cible de collaboration (médiane de 4) a été atteinte pour toutes les activités, avec très peu de variabilité dans les scores entre participants. Les éléments relatifs à la conduite d'activités faisant partie du processus de conception créative ainsi qu'à la négociation des choix de conception ont reçu le plus haut niveau de participation (moyenne de 3.66 et 3.60 resp.). Malgré un bon score, l'élément lié à la définition des priorités a tout de même obtenu le plus faible niveau de participation (moyenne = 3.20). Enfin, il est intéressant de noter que si aucun des participants n'a qualifié sa participation de passive sur l'un des items évalués, aucun d'entre eux n'a également indiqué le plus haut niveau de participation sur l'un des items. Cela peut s'expliquer par la nouveauté de l'expérience, mais aussi par le fait que les participants se sentent dépourvus en termes de parcours et d'expérience pour être préparés à concevoir par eux-mêmes des outils qu'ils n'ont utilisés jusqu'à présent que de manière accessoire.

Table 3: Résultats de l'étude de la qualité de participation de PaDLAD

Tâche principale	Moyenne (écart-type)	Médiane (min-max)
Planification et organisation	3.46 (0.64)	4 (2–4)
Processus de conception créative	3.66 (0.62)	4 (2–4)
Définition des priorités	3.20 (0.94)	4 (2–4)
Négociation	3.60 (0.82)	4 (2–5)
Réflexion et évaluation	3.40 (0.74)	4 (2–4)

Échelles : 1 = passif, 2 = partage d'informations, 3 = engagement et mobilisation, 4 = collaboration et 5 = autonomisation.

Expérience utilisateur de LADStudio. Nous avons calculé les résultats de l'UEQ suivant la procédure proposée par ses créateurs après avoir échelonné les réponses des participants de -3 (extrême négatif) à $+3$ (extrême positif) sur une échelle de Likert. Les scores entre $-0,8$ et $0,8$ reflètent une évaluation neutre, tandis que les scores supérieurs à $0,8$ indiquent une évaluation positive et ceux inférieurs à $0,8$ impliquent une évaluation négative.

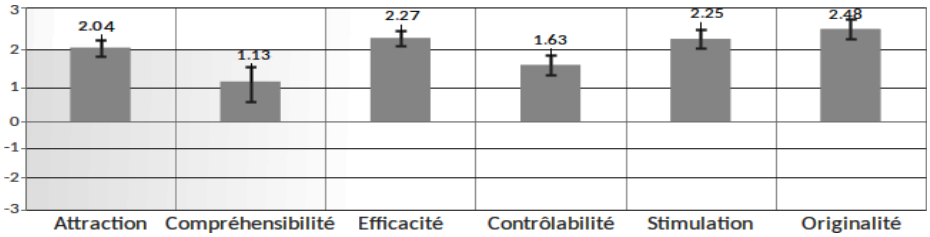


Fig. 4: Résultats de l'évaluation de l'expérience utilisateur de LADStudio.

Les notes obtenues pour LADStudio sont suffisamment élevées (Figure 4). Les scores des dimensions *Originalité*, *Efficacité* et *Stimulation* sont à un excellent niveau. Le score le plus élevé est celui de la dimension *Originalité*, avec une moyenne de $2,48$, suivi des dimensions *Efficacité* (moyenne $= 2,27$) et *Stimulation* (moyenne $= 2,25$). Ces scores sont d'un excellent niveau. La dimension *Contrôlabilité* a un bon score (moyenne $= 1,64$). La dimension *Compréhensibilité* a eu le résultat le plus bas (moyenne $= 1,13$), ce qui signifie que les participants ont éprouvé quelques difficultés à comprendre l'utilisation de l'outil.

Commentaires des participants. Les participants ont estimé que l'outillage de la phase idéation à travers PaDLAD fait émerger des idées de conception originales, notamment grâce à la dynamique d'échange et de partage. Il permet également de surmonter la difficulté de concevoir des TBA à partir d'une culture limitée en matière de données et de visualisation. La construction des vues en fonction des niveaux de conscience reflétait les étapes du raisonnement

et permettait une projection dans des scénarios d'utilisation réels, comme le dit un participant : *ce qui est intéressant, c'est l'accent mis sur l'histoire que vous voulez raconter à travers les cartes, ce qui est une perspective très intéressante, voire innovante*. Un autre participant a également soulevé les défis inhérents à la nécessité de construire les panneaux associés aux différents niveaux de conscience de la situation : *La possibilité de projeter le processus de raisonnement dans des représentations visuelle est attrayante, bien qu'elle ajoute de la complexité au fait de devoir déterminer les différentes étapes et de construire les écrans correspondants. De plus, parfois un seul écran suffit pour toutes les étapes.*

Dans les commentaires relatifs à LADStudio, les participants trouvaient que l'outil serait d'une grande utilité pour préciser les attentes et matérialiser les réponses, comme le souligne un participant : *un tel outil permet de ne pas rester au niveau idées et croquis, mais de continuer à travailler sur des prototypes fonctionnels*. Cependant, son utilisation requiert d'en comprendre le contexte et la raison d'être. En effet, les participants ont réussi à l'utiliser correctement avec de la pratique. Ils ont également apprécié l'accent mis sur la conception de TBA soutenant le processus de prise de sens de l'utilisateur, mais trouvaient comme pour PaDLAD que l'intégration de concepts théoriques relatifs au processus décisionnel pourrait rendre difficile l'appropriation initiale de l'outil.

5 Discussion & conclusion

La conception participative permet de créer des produits qui sont à la fois utiles et utilisables par les utilisateurs parce que conformes à leurs attentes. Dans cette contribution, nous avons proposé d'outiller ce processus pour les TBA afin de contribuer à remédier à leur adoption relativement limitée. Les résultats des aspects évalués montrent que les participants n'ont pas eu de difficulté dans l'utilisation des outils, montrant une attitude positive à l'égard de l'attrait de ces outils et de leur utilisabilité, ce qui est très motivant pour le développement futur de PaDLAD et LADStudio et qui répond à notre question de recherche *QR1*. L'étude SRLPS a démontré les capacités de PaDLAD à créer un environnement qui encourage la communication, l'échange et la créativité entre les participants (répondant à *QR2*). Enfin, l'expérience utilisateur offerte de LADStudio est d'un niveau acceptable dans l'ensemble et montre notamment que les participants considèrent l'outil comme très créatif, efficace et stimulant, ce qui nous a permis de répondre positivement à la question traiter *QR3*.

Malgré des résultats prometteurs, cette étude présente certaines limites. Tout d'abord, étant donné la nature qualitative de cette recherche, il est possible que des biais d'auto-sélection aient été introduits, car seuls les participants ayant un intérêt pour les TBA ont répondu à nos invitations. Comme l'ont indiqué les participants, l'adoption des TBA et plus généralement des outils LA en était à ses débuts, et leur mise en œuvre, tant par les participants que par leurs institutions, était donc le plus souvent d'ampleur limitée. En outre, l'étude ne peut fournir qu'une indication sur la qualité des outils en raison du nombre limité de participants. Afin de garantir la validité des conclusions, une évaluation plus

large doit être menée. La recherche a montré que l'utilisation du SUS fournit une mesure fiable de l'utilisabilité perçue d'un système, même avec un échantillon de taille relativement petite [22]. De plus, l'UEQ s'avère applicable même avec un groupe limité de participants [19]. Cependant, pour garantir la validité des résultats, une enquête plus approfondie impliquant un nombre plus important de participants serait nécessaire. De plus, une étude longitudinale à grande échelle sera essentielle pour évaluer la qualité des TBA créés et l'impact de l'utilisation de ces outils en se concentrant sur la perspective de la conception pédagogique.

Les études réalisées ont confirmé l'innovation de l'approche proposée ainsi que des outils développés pour résoudre deux problèmes majeurs qui entravent l'adaptation adéquate des TBA : le manque d'implication des utilisateurs finaux et la complexité technique de la création de TBA à partir de zéro. Des propositions innovantes pour l'adoption de TBA sont donc possibles lorsque les parties prenantes sont impliquées dans un environnement qui leur assure du soutien et de l'assistance. L'instrumentation de la conception de TBA permet d'envisager de nouvelles questions de recherche. Par exemple, les questions de *transférabilité* de TBA entre divers contextes d'utilisation, et d'invariants entre propositions issues de groupes d'utilisateurs différents forment des enjeux importants pour construire des TBA utiles au plus grand nombre.

Remerciements. Ce travail a été réalisé dans le cadre du projet AT41, financé par la DNE du Ministère de l'éducation nationale et le rectorat d'Orléans-Tours. Nous remercions également les personnes qui ont contribué à améliorer les outils, notamment : Sébastien Iksal, Katia Quelennec, Mathieu Vermeulen, Laurent Neyssenssas, Olivier Aubert, Rémi Venant, Kathy Casalino et Gabriel Trouvé.

References

1. Alhamadi, M.: Challenges, strategies and adaptations on interactive dashboards. In: Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization. pp. 368–371 (2020)
2. Alvarez, C.P., Martinez-Maldonado, R., Shum, S.B.: La-deck: A card-based learning analytics co-design tool. In: Proceedings of the tenth international conference on learning analytics & knowledge. pp. 63–72 (2020)
3. 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)
4. Brooke, J., et al.: Sus-a quick and dirty usability scale. Usability evaluation in industry **189**(194), 4–7 (1996)
5. Chatti, M.A., Muslim, A., Guliani, M., Guesmi, M.: The lava model: Learning analytics meets visual analytics. In: Adoption of data analytics in higher education learning and teaching, pp. 71–93. Springer (2020)
6. Dollinger, M., Liu, D., Arthars, N., Lodge, J.M.: Working together in learning analytics towards the co-creation of value. Journal of Learning Analytics **6**(2), 10–26 (2019)
7. Echeverria, V., Martinez-Maldonado, R., Granda, R., Chiluiza, K., Conati, C., Shum, S.B.: Driving data storytelling from learning design. In: Proceedings of the

8th International Conference on Learning Analytics & Knowledge. pp. 131–140. ACM (2018)
8. Gilliot, J.M., Iksal, S., Medou, D.M., Dabbebi, I.: Conception participative de tableaux de bord d'apprentissage. In: IHM'18: 30e Conférence Francophone sur l'Interaction Homme-Machine. pp. pp–119 (2018)
9. Gilliot, J.M., Sadallah, M.: A framework for co-designing effective lads supporting sensemaking and decision making. International Journal of Learning Technology (in press)
10. Hyett, N., Bagley, K., Iacono, T., McKinstry, C., Spong, J., Landry, O.: Evaluation of a codesign method used to support the inclusion of children with disability in mainstream schools. International Journal of Qualitative Methods **19**, 1–12 (2020)
11. Jivet, I., Scheffel, M., Drachsler, H., Specht, M.: Awareness is not enough: Pitfalls of learning analytics dashboards in the educational practice. In: Lavoué, É., Drachsler, H., Verbert, K., Broisin, J., Pérez-Sanagustín, M. (eds.) Data Driven Approaches in Digital Education. pp. 82–96. Springer, Cham (2017)
12. Keshavarzi, M., Bidgoli, A., Kellner, H.: V-dream: Immersive exploration of generative design solution space. In: International Conference on Human-Computer Interaction. pp. 477–494 (2020)
13. Laugwitz, B., Held, T., Schrepp, M.: Construction and evaluation of a user experience questionnaire. In: Symposium of the Austrian HCI and usability engineering group. pp. 63–76. Springer (2008)
14. Prieto-Alvarez, C.G., Martinez-Maldonado, R., Anderson, T.D.: Co-designing learning analytics tools with learners. In: Learning Analytics in the Classroom, pp. 93–110. Routledge (2018)
15. Sadallah, M., Gilliot, J.M.: Generating lads that make sense. In: Proceedings of the 15th International Conference on Computer Supported Education - Volume 1: CSEDU 2023. pp. 35–46. INSTICC, SciTePress (2023)
16. Sadallah, M., Gilliot, J.M., Iksal, S., Quelennec, K., Vermeulen, M., Neyssensas, L., Aubert, O., Venant, R.: Designing lads that promote sensemaking: A participatory tool. 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. pp. 587–593. Springer International Publishing, Cham (2022)
17. Santoso, H.B., Schrepp, M., Isal, R., Utomo, A.Y., Priyogi, B.: Measuring user experience of the student-centered e-learning environment. Journal of Educators Online **13**(1), 58–79 (2016)
18. Sarmiento, J.P., Wise, A.F.: Participatory and co-design of learning analytics: An initial review of the literature. In: Proceedings of the 12th International Learning Analytics and Knowledge Conference. pp. 535–541 (2022)
19. Schrepp, M., Hinderks, A., Thomaschewski, J.: Applying the user experience questionnaire (ueq) in different evaluation scenarios. In: International Conference of Design, User Experience, and Usability. pp. 383–392. Springer (2014)
20. 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 (2017)
21. Shaw, M.: The role of design spaces. IEEE software **29**(1), 46–50 (2012)
22. Tullis, T.S., Stetson, J.N.: A comparison of questionnaires for assessing website usability. In: Usability Professionals Association (UPA) 2004 Conference (2004)
23. Wang, F., Hannafin, M.J.: Design-based research and technology-enhanced learning environments. Educational Technology Research and Development **53**(4), 5–23 (2005)

Traçabilité *by design* : conception d'un système interactif pour améliorer la génération automatique de traces Git pendant une activité d'apprentissage

Mika Pons[1 [0000−0002−5844−8727]], Jean-Michel Bruel[2 [0000−0002−3653−0148]], Jean-Baptiste Raclet[3 [0000−0001−7357−912X]] et Franck Silvestre[1 [0000−0002−1134−8200]]

[1] IRIT, Université Toulouse Capitole, France
[2] IRIT, Université Toulouse 2 Jean Jaurès, France
[3] IRIT, Université Toulouse 3 Paul Sabatier, France

Résumé. Les *Learning Analytics* (LA) s'intéressent à la collecte et l'analyse des traces d'activités des apprenants dans le but de comprendre et améliorer leurs apprentissages. Dans cet article, nous nous intéressons aux traces laissées lors de l'utilisation de l'outil de gestion de versions Git. Les travaux existants sur le sujet présentent des limites quant à la qualité des traces qu'ils analysent : (1) la quantité et le contenu n'est pas toujours suffisant, (2) leur fiabilité relative peut entraîner une perte de données exploitables, et (3) la méthode de génération de ces traces n'est pas simple d'accès ou facilement applicable selon le public visé. Nous proposons un nouveau système interactif fondé sur Git et l'observation des modifications de fichiers afin de générer automatiquement des traces fiables et riches. Ce système sera expérimenté en contexte écologique prochainement et s'adresse à des contextes d'enseignement diversifiés.

Mots-clés : Learning Analytics, Traces d'apprentissage, Git, Système interactif

Abstract. *Learning Analytics* (LA) is collecting and analyzing traces of learners' activities in order to understand and improve learning. This paper focuses on traces left when using the version control system Git. Existing works on the topic have limitations regarding the quality of traces they analyze: (1) the quantity and content are not always sufficient, (2) their limited reliability can lead to a loss of exploitable data, and (3) the method of generating these traces is not easy to access or easily applicable depending on the target audience. We propose a new interactive system based on Git and the observation of file modifications to generate reliable, rich, and automatically collected traces. This interactive system will soon be experimented with in an ecological context soon and is intended for diversified teaching contexts.

Keywords: Learning analytics, Git, Traces, Interactive system.

1 Introduction

La première conférence sur le champ des *Learning Analytics* le définit comme "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 produit" [1].

Si l'utilisation d'outils de gestion de versions est très répandue dans l'industrie du logiciel, on en retrouve également une exploitation dans un contexte éducatif, essentiellement dans l'apprentissage de l'informatique [5,4]. L'intégration de tels outils dans les enseignements permet non seulement d'accompagner la professionnalisation des étudiants, mais aussi de faciliter le travail collaboratif au sein d'un groupe d'étudiants. Ils constituent enfin un moyen simple d'obtenir des traces de l'activité des étudiants grâce aux actions de *commits* réalisées au cours de la progression des activités afin de versionner le travail.

À la croisée des *Learning Analytics* et des logiciels de gestion de versions, on voit émerger des travaux fondés sur l'analyse de traces produites avec l'utilisation de logiciels de gestion de versions dans l'apprentissage de l'informatique [10]. Certains travaux exploitent des techniques de *Process Mining* pour l'analyse des traces en extrayant des métriques liées au comportement d'apprentissage des étudiants [6,2].

Cependant, ces travaux présentent un certain nombre de limites concernant les traces générées grâce aux logiciels de gestion de versions tel que Git. En effet, le contenu de ces traces n'est pas suffisant pour certaines analyses nécessitant de connaître des informations temporelles plus précises. Aussi, ces traces peuvent être d'une fiabilité relative quand leur génération repose sur un processus déclaratif et ouvert de l'étudiant. Enfin, ces traces sont générées avec des méthodes que l'on peut difficilement appliquer avec des étudiants inexpérimenté avec Git.

Cela nous amène à nous demander quel système interactif de génération de traces à destination des étudiants concevoir pour générer des traces Git permettant de lever ces limites, et plus précisément :

RQ1 : Comment augmenter la quantité et le contenu de traces Git à l'aide d'un système interactif ?

RQ2 : Comment augmenter la fiabilité des traces Git à l'aide d'un système interactif ?

RQ3 : Comment un système interactif peut-il simplifier le processus de génération des traces Git pour une audience qui n'est pas familière avec Git ?

Dans un premier temps, nous ferons un tour d'horizon des travaux qui utilisent des traces Git pour améliorer l'enseignement en section 2, tout en mettant en lumière leurs limites. Dans un deuxième temps, en section 3, nous présenterons le système interactif *LAWG* conçu pour répondre aux problématiques soulevées par nos questions de recherche. Enfin, nous expliquerons comment les fonctionnalités de *LAWG* y répondent dans la section 4.

2 État de l'art

Plusieurs articles [3,5] ont fait l'expérience d'intégrer un logiciel de gestion de versions dans l'apprentissage de l'informatique et en montrent les bénéfices. [11] et [4] montrent plus précisément l'intérêt des logiciels de gestion de versions décentralisés, comme Mercurial et Git.

Dans le domaine du *Process Mining* (PM), plusieurs travaux [2,7,13] s'intéressent à analyser les traces Git générées dans le contexte de cours de développement logiciel. En 2021, [6] proposent une méthode pour analyser des traces Git de projets d'étudiants à l'aide du PM. Ils s'intéressent à 13 activités correspondant à des caractéristiques d'un *commit*. Plus précisément, est regardé s'il s'agit d'un commit contenant des additions ou des suppressions de lignes, s'il s'agit d'un gros ou d'un petit *commit*, d'un *commit* lié à la gestion des branches (nouvelle branche, *merge*, *pull request*), ou encore s'il s'agit d'un *commit* pour créer ou modifier des fichiers de test. Cet article montre qu'il est possible d'identifier des comportements d'apprentissages des étudiants en utilisant la technique du PM. En revanche, seules les informations fournies par Git dans un *commit* sont considérées pour être analysées. Les traces Git ne contiennent pas d'informations contextuelles quant à la progression d'un étudiant dans une activité pédagogique, comme la résolution d'une question bien identifiée. Il n'y a pas d'informations temporelles concernant les dates de début d'activité de résolution d'une question et donc n'adresse pas (RQ1). Enfin, cette méthode d'analyse ne traite pas les questions de la fiabilité de la génération des traces et de sa simplicité d'accès, comme visé dans (RQ2, RQ3).

En 2018, [14] développent un protocole d'apprentissage fondée sur la revue de code et le développement guidé par les tests. En pratique, les étudiants suivent une feuille d'exercices et doivent marquer la résolution d'une question en effectuant des *commits* sur leur dépôt Git associé. Des corrections et des revues de codes sont planifiées en tenant compte de la progression du groupe. Pour aider à l'orchestration de ces différentes phases, un tableau de bord, *Git4School* [9], est utilisé. À partir des *commits* réalisés par les étudiants et des jalons (revues de code, corrections) renseignés par l'enseignant, ce tableau de bord permet de visualiser un certain nombre d'indicateurs quant à la progression des étudiants par rapport à leurs camarades. Il permet également d'exporter les traces enrichies par le contexte de résolution des questions.

Nous avons analysé ces traces avec des techniques de PM [8] pour extraire des indicateurs sur le comportement d'apprentissage des étudiants. Malheureusement, les résultats étaient limitées par le manque d'informations contenues dans ces traces. En effet, il ne nous était pas possible d'identifier précisément le début d'une activité car le *commit* ne donnait que l'achèvement de l'activité concernant une question, n'adressant donc pas (RQ1). De plus, l'exercice terminé est identifié grâce à un message prédéfini associé au *commit*. Ce processus déclaratif et ouvert (l'étudiant peut entrer le message qu'il souhaite pour le commit) a conduit à des traces parfois inutilisables à cause d'erreurs dans le message ou d'oublis des étudiants, ce qui soulève l'enjeu de la fiabilité des traces mentionné dans (RQ2). Aussi, l'approche de génération manuelle des traces par

les étudiants avec Git pose un problème de simplicité d'accès et d'applicabilité de l'approche dans des contextes autres (RQ3).

Une étude de 2022 [12] explore un moyen alternatif de générer automatiquement et de visualiser des traces d'activité des étudiants lors de séances de TP. La plateforme de visualisation des données EnCourse présente des indicateurs bruts directement liés aux *commits* faits par les étudiants, comme le temps passé sur le projet, la fréquence des *commits*, ou le nombre de lignes ajoutées ou supprimées. Elle fournit également des indicateurs interprétés comme la détection de cas de "malhonnêteté académique" ou des alertes lorsque les étudiants sont en retard. Pour automatiser la génération des traces, [12] ont fait le choix d'intégrer les commandes de commit dans un fichier Makefile. Ainsi, un commit est fait chaque fois que le Makefile est exécuté, à chaque compilation.

Nous voyons trois limites concernant les traces générées via ce fonctionnement. Premièrement, une compilation n'est pas un événement fréquent et le grain de traçage est donc un peu grossier (RQ1). Deuxièmement, l'usage d'un fichier Makefile est une contrainte qui rend difficile la généralisation à d'autres domaines que celui de l'informatique (RQ3). En effet, et ce même en informatique, la compilation est une action spécifique à la programmation. Un cours de génie logiciel où les étudiants ont pour objectif, par exemple, de rédiger un dossier de conception ne nécessitera pas de compilation, et l'intégration d'un Makefile sera impossible. Enfin, pour calculer le temps passé en activité, la plateforme compte, à partir des *commits*, des fenêtres d'activité fixes dont la durée est choisie arbitrairement. Or rien ne permet de fixer a priori une telle durée d'activité de l'étudiant.

En conclusion de cet état de l'art, à notre connaissance, il n'existe pas de travaux sur un système interactif permettant de générer des traces qui lèvent les limites de contenu, de fiabilité et de simplicité que nous venons d'identifier. Pour ce faire, nous avons cherché un compromis entre *Git4School* et EnCourse. Nous avons donc conçu un système interactif de génération automatique de traces qui prend en compte le besoin d'avoir des informations contextuelles sur la progression des étudiants en temps réel, tout en maximisant la fiabilité des traces et la simplicité d'accès de leur méthode de production.

3 Le système interactif *LAWG* pour la génération automatique de traces Git d'activité

Nous nous plaçons dans le contexte de résolution de feuilles d'exercices[4] données lors de séances de travaux pratiques (TP). Ces exercices peuvent être constitués de plusieurs *questions*. Nous définissons la notion de *session de travail* comme étant une période d'activité d'un étudiant, durant laquelle il manipule des ressources (fichiers de programmation, textes, images, ...), qui constituent son *espace de travail*, pour résoudre des questions. Un étudiant peut travailler en dehors

[4]Un exemple de feuille d'exercices est donné ici : https://gist.github.com/raclet/c05729a4ae78082365ed0a1f8226fe55

```yaml
repo_path: .
ssh_path: /home/user/.ssh/id_rsa
questions: []
groups: []
```

Listing 1 : Structure du fichier `settings.yml`

des séances de TP encadrées, une session de travail ne se limite donc pas aux seules séances de TP.

L'espace de travail d'un étudiant existe localement sur sa machine, et à distance. L'espace à distance permet à l'enseignant d'accéder au travail de l'étudiant, et à ce dernier de synchroniser son espace local s'il doit travailler sur plusieurs machines ou avec d'autres étudiants. Lorsqu'un étudiant a fini de travailler sur une question, il valide ses changements, puis passe à la question suivante. Dans notre cas, l'espace de travail est géré avec le logiciel de gestion de versions Git[1]. Pour valider ses changements, il effectue un *commit*, ce qui enregistre localement toutes les modifications faites aux fichiers depuis son dernier commit. L'enregistrement est toujours associé à un message entré par l'étudiant au moment du commit (par exemple, `Exercice 1 résolu`). Pour synchroniser l'espace de travail distant, son *dépôt*, avec son espace local, il exécute la commande Git *push*, et pour synchroniser son espace local avec celui à distance, il exécute la commande Git *pull*. Enfin, l'enchaînement des commits sur l'espace de travail constitue une *branche*. Il existe une branche sur laquelle l'étudiant évolue par défaut, mais il peut faire le choix d'en créer une nouvelle à partir de celle-ci. Les branches évoluent de façon distincte, de sorte que l'étudiant peut changer de branche pour isoler certains types de modifications (travail sur une fonctionnalité, ...).

Pour répondre aux questions de recherche posées dans ce papier, nous avons conçu un système interactif écrit en Python et open-source[5]. Pour que tous les étudiants puissent l'utiliser, il est disponible sous forme de fichier exécutable pour les trois principaux systèmes d'exploitation : Windows, MacOS et Linux.

En agissant comme une interface entre l'étudiant et son dépôt Git, *LAWG* génère des traces enrichies et fiables tout en abstrayant Git pour les étudiants.

3.1 Configuration

Une configuration minimale est nécessaire, de la part de l'enseignant dans la phase de conception du sujet de TP, et de la part des étudiants en phase d'exécution du sujet.

[1] https://git-scm.com/

[5] Disponible sur Github à cette adresse : https://github.com/git4school/LAWG.

Pour préparer le sujet, l'enseignant met en place un espace de travail "template", composé d'un ensemble requis d'éléments permettant aux étudiants de travailler. Dans cet espace de travail, il doit intégrer et configurer *LAWG* (les fichiers exécutables correspondant aux systèmes d'exploitations que ses étudiants sont susceptibles d'utiliser). L'enseignant doit créer le fichier de configuration `settings.yml`, dont les champs requis sont listés dans l'extrait 1. Nous voulons que l'utilisation du système soit la plus simple et fiable possible. Pour cela, il est nécessaire de lui fournir la liste des questions auxquelles l'étudiant devra répondre (le champs `questions`). C'est également à cette étape que les champs `groups` et `repo_path` peuvent être renseigner. Une fois cela fait, l'enseignant peut publier le sujet et le dépôt "template".

C'est au tour de l'étudiant de créer son espace de travail à partir du dépôt "template" et de terminer la configuration. Il ne reste que le champs `ssh_path` dans lequel il faut indiquer le chemin vers la clé SSH enregistrée dans la plateforme qui héberge le dépôt Git distant. Une des principales problématiques avec la conception d'un tableau de bord, comme *Git4School*, est l'identification des étudiants. Pour ce faire, il était demandé aux étudiants de remplir le fichier README de leur projet avec leur nom et leur groupe. Le README pouvait contenir d'autres informations comme des instructions, et le tableau de bord devait donc extraire l'identité à l'aide d'un *parser*. De fait, le succès de l'identification de l'étudiant était fortement dépendant de celui-ci, qui devait (1) ne pas oublier de remplir le README, (2) remplir le README correctement, et (3) ne pas modifier le README ultérieurement de sorte que le *parsing* ne peut plus se faire correctement. Afin de fiabiliser ce processus, au premier démarrage, *LAWG* demande à l'étudiant d'entrer son nom et son groupe et génère le fichier IDENTITY.json. Ce fichier sera ensuite lu par *Git4School* pour identifier chaque étudiant.

Notre système interactif offre 3 grandes fonctionnalités que nous explicitons dans les sous-sections suivantes :

- Encadrer les sessions de travail, sauvegardant toutes les ressources à la fin de chaque session ;
- Observer toutes les modifications apportées aux fichiers de l'espace de travail pour les sauvegarder en temps réel ;
- Offrir une interface en ligne de commande aux étudiants pour leur permettre de marquer une question comme résolue.

3.2 Encadrement des sessions de travail

Le début et la fin d'une session de travail sont jalonnés par des commits nommés `"Resume"` et `"Pause"` dans la branche courante. Comme illustré dans la figure 1, nous avons fait le choix d'utiliser une autre branche, `auto`, pour isoler l'exécution des commits automatiques. Cette branche est créée à partir de la branche courante au lancement de *LAWG* si elle n'existe pas. Cela permet à

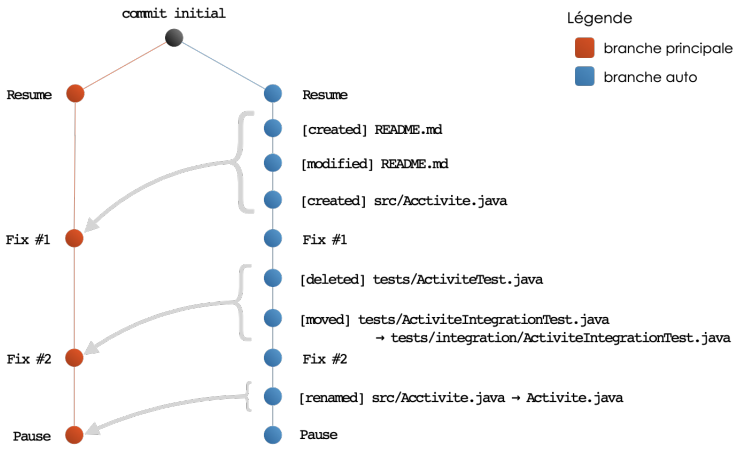


FIG. 1 : Exemple de l'utilisation de *LAWG* lors de la réalisation d'exercices de travaux pratiques

l'étudiant d'avoir une branche courante simple avec peu de commits, et à l'enseignant d'avoir une branche qui retrace l'activité complète de l'étudiant. Certain commits sont présents dans les deux branches.

Lors de la fin d'une session de travail, marquée par l'utilisation de la commande `exit`, tous les changements en cours sont sauvegardés sous un commit `"Pause"` et envoyés dans l'espace de travail distant. Ensuite, tous les fichiers de son espace de travail local sont supprimés. De cette façon, aux yeux de l'étudiant, son espace de travail est vide, il ne reste plus que l'exécutable du système et le fichier de configuration. L'étudiant doit lancer le système interactif pour restaurer son espace de travail. Le système peut ainsi tracer automatiquement le démarrage d'une nouvelle session de travail, représenté par un commit `"Resume"` dans la branche `auto`.

3.3 Génération automatique des traces d'activité

Dès le lancement de *LAWG*, les modifications des fichiers de l'espace de travail sont observées et donnent lieu à des événements système. Chaque événement entraîne la production d'un commit dans la branche `auto`. Les événements observés sont : la *modification*, la *création*, la *suppression*, le *déplacement* et le *renommage* d'un fichier.

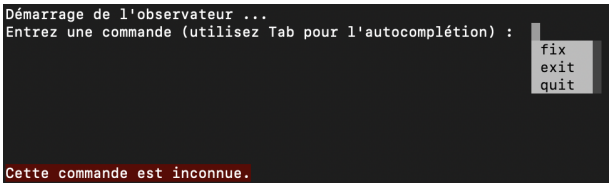
Pour ces commits automatiques, nous avons défini une forme de message minimaliste pour faciliter le traitement par la suite :

- `"[moved] <chemin_fichier> -> <chemin_nouveau_dossier>"`
- `"[renamed] <chemin_fichier> -> <nouveau_nom_fichier>"`
- `"[<evenement>] <chemin_fichier>"`

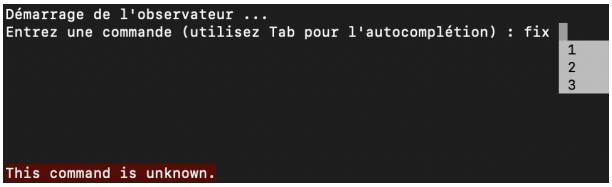
Ainsi, dans l'exemple de la figure 1, on peut voir que l'étudiant démarre sa session de travail avec la création, puis la modification d'un fichier README, et crée un fichier `Acctivite.java` avant de marquer la question `#1` comme résolue. Il fait ensuite des opérations sur des fichiers de tests, résout la question `#2`, puis se rend compte qu'il a fait une faute sur le nom d'un fichier et corrige l'erreur. L'étudiant ferme ensuite sa session, ce qui a pour effet de produire le commit `"Pause"`.

3.4 Interface de résolution d'un exercice

Notre système interactif propose également une interface en ligne de commande, qui met notamment à disposition la commande `fix`. Elle permet à l'étudiant de marquer une question comme résolue, de sauvegarder les changements associés et de les envoyer dans l'espace de travail distant. Plus précisément, tous les changements depuis le dernier appel à cette même commande, ou depuis le lancement d'une session de travail, sont validés sous forme d'un commit dans la branche courante et dans la branche `auto` (dont le message est de la forme `"Fix <question>"`) puis envoyés dans l'espace de travail distant. Dans l'exemple donné en figure 1, l'étudiant a utilisé la commande `fix` pour les questions `#1` et `#2`, dont les commits associés sont visibles dans les deux branches, puis la commande `exit`, qui a produit le commit `"Pause"`.



(a) Suggestion des commandes



(b) Suggestion des questions

FIG. 2 : Captures d'écran des suggestions de *LAWG*

Le système offre une suggestion des commandes disponibles (cf. figure 2a) en fonction des caractères déjà entrés et l'étudiant ne peut pas lancer une commande qui n'existe pas ou qui est mal formée. Lorsqu'il reconnaît la commande `fix`, le système interactif affiche la liste des questions auxquelles l'étudiant doit répondre (cf. figure 2b), renseignées auparavant lors de la configuration. Le système a

été conçu de manière à pouvoir spécifier et implanter des commandes selon le contexte d'enseignement facilement.

3.5 Limites du système interactif

LAWG a des limites qu'il faut prendre en compte avant toute intégration.

L'étudiant doit encore effectuer plusieurs actions manuellement. En effet, il doit enregistrer sa clé SSH publique sur Github lorsque son espace de travail y est hébergé, cloner son dépôt pour la première fois et utiliser Git pour faire un pull afin éventuellement de synchroniser l'espace de travail entre deux machines utilisées. La nécessité d'avoir recours à une clé SSH est un frein à lever pour lequel nous étudions des pistes simplificatrices car nous visons à terme un public non informaticien.

De plus, *LAWG* ne récupère pas les changements récents faits sur le dépôt distant via la commande *pull* car il ne pourrait pas gérer les éventuels conflits de manière automatique. Cela peut poser des problèmes lorsque plusieurs étudiants travaillent sur un même projet ou si un étudiant travaille sur plusieurs machines différentes.

Dans le cas où les étudiants sont amenés à utiliser plusieurs branches, l'utilisation de *LAWG* est pour le moment déconseillée. Il ne gère pas correctement le changement de branche. Si le contenu de l'espace de travail change entre deux branches, le système va sauvegarder tous les fichiers différents de la branche de départ.

Il subsiste une dépendance au déclaratif de la part de l'étudiant, c'est-à-dire que l'étudiant doit déclarer explicitement les questions qu'il résout petit à petit. Il n'y a pas d'évaluation objective et automatique du travail de l'étudiant.

4 Quelles sont les bénéfices de l'utilisation du système interactif ?

Dans cette section, nous expliquons comment les fonctionnalités de *LAWG* permettent de répondre aux trois questions de recherche énoncées dans la section 1.

4.1 RQ1 - Contenu

La génération des traces pour chaque modification de fichier (cf. sous-section 3.3) permet de tracer finement l'activité d'un étudiant pendant une session de travail. En effet, ce grain plus fin de trace permet d'avoir le premier signe d'activité concernant une question *n* après la résolution d'une question *n-1*. Ainsi, nous pouvons nous approcher plus précisément de la durée réelle d'activité de l'étudiant. En générant une trace pour chaque modification de fichier, il est possible d'identifier des informations sur le comportement des étudiants via les noms de ces fichiers. Un exemple concret, dans le cas de la figure 3, est l'identification du bon respect du Test-Driven-Development (TDD). Cette figure est un extrait

44	fbf16ec	Mon, 10 Apr 2023 18:28:55 +0200	Pause
45	ea3f9d1	Mon, 10 Apr 2023 18:26:46 +0200	Fix #2b
46	19c7009	Mon, 10 Apr 2023 18:26:06 +0200	[modified] src/test/java/doremi/doremi/BandTest.java
47	6890448	Mon, 10 Apr 2023 18:23:36 +0200	[modified] src/main/java/doremi/domain/Band.java
48	89a6b70	Mon, 10 Apr 2023 18:23:31 +0200	[modified] src/main/java/doremi/domain/Band.java
49	1157e96	Mon, 10 Apr 2023 18:23:15 +0200	[modified] src/test/java/doremi/doremi/BandTest.java
50	fa4054a	Mon, 10 Apr 2023 18:23:13 +0200	[created] src/test/java/doremi/doremi/BandTest.java
51	fa0c6fa	Mon, 10 Apr 2023 18:22:39 +0200	[modified] src/main/java/doremi/domain/Album.java
52	ce3c3bd	Mon, 10 Apr 2023 18:22:23 +0200	[modified] src/main/java/doremi/domain/Album.java
53	1c5090f	Mon, 10 Apr 2023 18:21:47 +0200	[modified] src/test/java/doremi/AlbumTest.java
54	2312a58	Mon, 10 Apr 2023 18:21:34 +0200	[modified] src/test/java/doremi/AlbumTest.java
55	533580e	Mon, 10 Apr 2023 18:21:33 +0200	[created] src/test/java/doremi/AlbumTest.java
56	ce3dd1d	Mon, 10 Apr 2023 18:19:26 +0200	Fix #2a
57	be88b27	Mon, 10 Apr 2023 18:19:11 +0200	[modified] src/main/java/doremi/domain/Article.java
58	6e9899f	Mon, 10 Apr 2023 18:19:07 +0200	[modified] src/main/java/doremi/domain/Article.java
59	6b853e4	Mon, 10 Apr 2023 18:18:21 +0200	[modified] src/main/java/doremi/domain/Article.java
60	6da0f26	Mon, 10 Apr 2023 18:17:59 +0200	[modified] src/main/java/doremi/domain/Article.java
61	ecb0a7a	Mon, 10 Apr 2023 18:17:29 +0200	[modified] src/test/java/doremi/ArticleTest.java
62	151e5c2	Mon, 10 Apr 2023 18:17:20 +0200	[created] src/test/java/doremi/ArticleTest.java
63	c13b362	Mon, 10 Apr 2023 18:16:09 +0200	Fix #1
64	fe51014	Mon, 10 Apr 2023 18:15:52 +0200	[modified] src/main/java/doremi/controllers/IndexController.java
65	e42f65f	Mon, 10 Apr 2023 18:14:54 +0200	[modified] src/test/java/doremi/IndexControllerTest.java
66	81bfb45	Mon, 10 Apr 2023 18:14:47 +0200	[created] src/test/java/doremi/IndexControllerTest.java
67	55f182b	Mon, 10 Apr 2023 18:14:22 +0200	[modified] src/main/java/doremi/controllers/IndexController.java
68	9f3e898	Mon, 10 Apr 2023 18:14:05 +0200	[modified] src/main/java/doremi/controllers/IndexController.java
69	424e099	Mon, 10 Apr 2023 18:13:43 +0200	[modified] src/main/java/doremi/controllers/IndexController.java
70	dff1b1d	Mon, 10 Apr 2023 18:13:28 +0200	[created] src/main/java/doremi/controllers/IndexController.java
71	3191ec0	Mon, 10 Apr 2023 18:12:33 +0200	[modified] src/main/resources/templates/index.html
72	9c8158d	Mon, 10 Apr 2023 18:12:24 +0200	[modified] src/main/resources/templates/index.html
73	79bbdbd	Mon, 10 Apr 2023 18:12:17 +0200	[created] src/main/resources/templates/index.html
74	8a0b7c7	Mon, 10 Apr 2023 18:10:07 +0200	Resume

FIG. 3 : Extrait des traces générées avec *LAWG*

des traces que nous avons générées pour tester le système[6]. Elles simulent un étudiant utilisant le système pour répondre à des questions sur une feuille de travail d'un cours réel. Chaque ligne en rouge correspond à une trace générée par *LAWG*. Nous pouvons voir que le volume des traces est significativement augmenté et couvre une plage de temps plus étendue que sans l'utilisation du système (seulement les commits nommés `"Fix ..."`). Nous pouvons voir aux lignes 62 et 61 que l'étudiant intègre le test `ArticleTest`, puis modifie la classe associée `Article` grâce à la convention de nommage des tests. Ceci est donc représentatif du TDD. D'autre part, nous pouvons voir que l'étudiant modifie les classes `IndexController` à la ligne 67 avant le test associé `IndexControllerTest` à la ligne 66. On voit ici que le TDD n'est pas respecté, et comparer les noms de fichiers permettrait d'automatiser cette identification à partir de ces traces.

De plus, la gestion des sessions de travail (cf. sous-section 3.2) permet d'intégrer le début et la fin de ces sessions dans les traces. Aussi, le mécanisme associé qui vide et restaure l'espace de travail, permet de nous assurer de l'utilisation

[6]Les données sont disponibles en accès libre sur https://osf.io/t3wqr/?view_only=69907570f39046edba382a5e855ed26a

systématique du système interactif pour chaque session de travail, et donc de l'exhaustivité des traces.

4.2 RQ2 - Fiabilité

Avec l'interface de résolution de question (cf. sous-section 3.4), nous sommes passés d'une génération de traces déclarative et ouverte à une génération de traces déclarative mais fermée. En effet, la commande `fix` génère une trace dont le message est normalisé. La validation de la commande empêche l'étudiant de marquer comme résolue une question qui n'est pas dans la liste donnée à la configuration. On évite de cette façon les erreurs de la part des étudiants. Grâce à son automatisation en partie (cf. sous-section 3.3), la génération de traces devient semi-automatique et fermée. Une grande partie des traces générées par *LAWG* n'est plus dépendante de la déclaration de la part de l'étudiant, ce qui réduit la possibilité d'oublis.

4.3 RQ3 - Simplicité d'accès

Le système interactif permet de s'intégrer plus facilement dans des domaines non informatiques. Il abstrait l'utilisation de Git pour la génération des traces, et peut donc être utilisé pour une audience avec un profil en informatique mais qui n'est pas familiarisé ave Git. Le déclenchement de la génération automatique des traces (cf. sous-section 3.3) lors de la simple modification d'un fichier est un pas en faveur de l'utilisation du système dans n'importe quel domaine nécessitant de travailler sur ordinateur, notamment quand les contributions attendues sont des productions de textes.

5 Conclusion

Après avoir identifié des limites des outils qui génèrent des traces à travers des dépôts Git, nous avons présenté le système interactif que nous avons conçu pour les lever, appelé *LAWG*. Notre système interactif permet la génération automatique de traces à chaque modification de fichier par l'étudiant. Il fournit également une commande permettant à l'étudiant de valider la résolution d'une question. Nous contribuons ainsi à abstraire la production de traces d'activité, et à améliorer leur fiabilité et leur contenu. Notre système interactif sera expérimenté cette année dans des cours d'apprentissage de l'informatique pour produire de nouvelles traces. Nous avons établi une approche qui sépare l'accès aux données personnelles du travail de recherche. Seuls les enseignants ont accès aux données personnelles via le tableau de bord *Git4School* et mettent ces données à la disposition des chercheurs une fois anonymisées (au moyen d'un script que nous fournissons). Pour faciliter cette démarche et assurer la conformité au RGPD par conception, nous prévoyons d'intégrer l'anonymisation des données dans l'exportation à partir de *Git4School*.

Références

1. LAK '11 : Proceedings of the 1st International Conference on Learning Analytics and Knowledge (2011)
2. Eskofier, B.M. : Exploration of process mining opportunities in educational software engineering-the gitlab analyser. In : Proc. of The 13th International Conference on Educational Data Mining (EDM 2020) (2020)
3. Glassy, L. : Using version control to observe student software development processes. Journal of Computing Sciences in Colleges (2006)
4. Laadan, O., Nieh, J., Viennot, N. : Teaching operating systems using virtual appliances and distributed version control. In : Proceedings of the 41st ACM technical symposium on Computer science education (2010)
5. Lawrance, J., Jung, S., Wiseman, C. : Git on the cloud in the classroom. Proceeding of the 44th ACM technical symposium on Computer science education - SIGCSE '13 (2013)
6. Macak, M., Kruzelova, D., Chren, S., Buhnova, B. : Using process mining for Git log analysis of projects in a software development course. Education and Information Technologies (2021)
7. Mittal, M., Sureka, A. : Process mining software repositories from student projects in an undergraduate software engineering course. In : Companion proceedings of the 36th international conference on software engineering (2014)
8. Pons, M., Bruel, J.M., Raclet, J.B., Silvestre, F. : Finding behavioral indicators from contextualized commits in software engineering courses with process mining. In : Frontiers In Software Engineering Education (2023)
9. Raclet, J.B., Silvestre, F. : Git4School : A dashboard for supporting teacher interventions in software engineering courses. In : European Conference on Technology Enhanced Learning (2020)
10. Robles, G., González-Barahona, J.M. : Mining student repositories to gain learning analytics. an experience report. In : 2013 IEEE Global Engineering Education Conference (EDUCON) (2013)
11. Rocco, D., Lloyd, W. : Distributed version control in the classroom. In : Proceedings of the 42nd ACM technical symposium on Computer science education (2011)
12. Rodriguez-Rivera, G., Turkstra, J., Buckmaster, J., LeClainche, K., Montgomery, S., Reed, W., Sullivan, R., Lee, J. : Tracking large class projects in real-time using fine-grained source control. In : Proceedings of the 53rd ACM Technical Symposium on Computer Science Education V. 1 (2022)
13. Shynkarenko, V., Zhevaho, O. : Application of constructive modeling and process mining approaches to the study of source code development in software engineering courses. Journal of Communications Software and Systems (2021)
14. Silvestre, F., Raclet, J.B. : Développement dirigé par les tests et revue de code par les pairs pour l'apprentissage de la programmation. In : Ludovia CH : 1ère édition sur le thème" Émanciper l'école et la société avec le numérique ?" (2018)

Co.LAB, une plateforme d'accompagnement pour la conception et l'évaluation collaboratives de jeux destinés à un usage éducatif

Éric Sanchez[1], Estelle Prior[1], Nadine Mandran[2], Sandra Monnier[3], Audrey Huguenin[3], Maxence Laurent[3], Mariem Jaouadi[1], Dominique Jaccard[3]

[1]Université de Genève, LIP/TECFA Genève, Suisse
prénom.nom@unige.ch
[2]Université de Grenoble Alpes, LIG Grenoble, France
nadine.mandran@univ-grenoble-alpes.fr
[3]Haute Ecole d'Ingénierie et de Gestion du Canton de Vaud, MIE ALBASIM, Yverdon, Suisse
prénom.nom@heig-vd.ch

Résumé. Nous présentons ici la plateforme co.LAB et les résultats de l'évaluation de ses apports comme soutien à conception et l'évaluation de jeux destinés à un usage pédagogique. L'évaluation se base sur la littérature du domaine, l'expérience acquise dans le cadre de projets de recherche sur la conception de jeux éducatifs une expérimentation suivie de la complétion d'un questionnaire et d'un *focus group* avec les utilisateurs. Cela nous permet de présenter et de discuter les apports et les limites de la plateforme.

Mots-clés : Recherche orientée par la conception, Game design collaboratif

Abstract. We present here the co.LAB platform and the results of the evaluation of its contributions to support the design and assessment of learning games. The evaluation is based on the literature review, experience gained from research projects on learning game design and an experimentation followed by the completion of a questionnaire and a focus group with users. This allows us to present and discuss the contributions and limitations of the platform.

Keywords : Design-Based Research, Collaborative Game Design

1 Introduction

Les intérêts des praticiens et des chercheurs se rejoignent parfois dans le cadre de projets de recherche collaboratifs orientée par la conception (ROC) (*design-based research*) [1] qui permettent d'atteindre simultanément des objectifs pragmatiques (l'amélioration de l'apprentissage) et théoriques (l'étude des principes de conception). Néanmoins, il est difficile de mettre en place un projet de ROC. Parmi les difficultés rencontrées, il faut relever la complexité et la durée du processus. Ainsi, notre objectif consiste à développer une méthode pour accompagner la co-conception et la co-

évaluation de jeux éducatifs. Il se traduit, en particulier, dans la conception d'une plateforme permettant de formaliser et d'implémenter cette méthode. Dans le présent article nous présentons une ROC en tant que processus de co-conception et co-évaluation de jeux et des points de vigilance à prendre en compte pour son implémentation. Nous présentons ensuite la plateforme co.LAB destinée à soutenir ce processus au regard de sa capacité à accompagner l'équipe en charge du projet (évaluation par inspection) et une expérimentation qui a permis d'évaluer son utilisabilité.

2 Conception et évaluation de jeux éducatifs

La conception et l'évaluation sont au cœur de la recherche orientée par la conception. La conception est un processus itératif monitoré par une évaluation continue [2]. Le modèle que nous avons retenu pour le représenter se présente sous la forme de deux cycles imbriqués (Figure 1) qui décrivent les différentes étapes d'une ROC. Le cycle de conception, inspiré du modèle ADDIE [3], comprend une phase d'analyse du contexte et de définition des objectifs pédagogiques (*Analyse*) ; une phase de design du jeu ainsi que du scénario d'usage (*Design*) ; une phase de réalisation du jeu (*Développement*) ; une dernière phase d'expérimentation du jeu dans le contexte auquel il est destiné (*Implémentation*). Ces phases s'appuient sur des modèles ou concepts théoriques relatifs aux relations entre jeu et apprentissage.

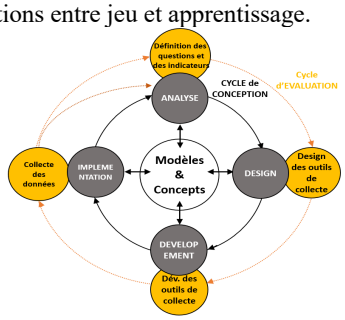


Fig. 1. Recherche orientée par la conception

Le cycle d'évaluation, conduit en parallèle avec le précédent, est initié lorsque chercheurs et praticiens s'accordent sur un problème pédagogique à résoudre, la formulation de questions de recherche et la définition d'indicateurs permettant de les traiter. Le cycle se poursuit avec la conception et le développement des outils de collecte des données. Ces données sont produites lors de l'expérimentation du jeu. Leur traitement et analyse permettent de boucler l'itération et d'initier de nouveaux cycles. Ces travaux permettent d'une part la mise en place d'une solution éducative répondant à un problème de terrain, et d'autre part, la mise à l'épreuve du modèle théorique.

La ROC est un processus multi-expertises qui nécessite la collaboration de chercheurs, de praticiens, d'experts du domaine enseigné, de graphistes, de développeurs et de *game designers*. Par ailleurs, Les méthodes d'ingénierie des jeux sont multiples et nous reprenons ici les résultats d'un état de l'art [4] qui a permis d'en dégager les principes fondamentaux. Il s'agit de s'assurer que l'ensemble des membres

de l'équipe connaissent la méthode retenue et puisse l'adapter au contexte du projet. Cela suppose une vision commune des objectifs et du processus [5] ainsi que des modèles théoriques mobilisés [6]. Cela suppose également l'utilisation de méthodes et d'outils spécifiques [7]. La méthode utilisée doit permettre le partage des savoirs entre les différents experts [8] et la spécification des compétences et rôles de ces experts ainsi qu'une planification et une répartition des tâches. Ces tâches doivent être spécifiées, les livrables attendus et les décisions explicitées, mais des réajustements et des renégociations doivent rester possibles [9]. La prise en compte du contexte et la collaboration avec les praticiens favorise la mise en place d'approches centrées utilisateur. Enfin, la méthode doit intégrer une réflexion éthique qui, en raison de son caractère évolutif, peut être qualifiée d'éthique située [10].

Cette revue de la littérature permet également d'identifier des questions relatives à la problématisation de la question éducative qui implique que les partenaires d'un projet de recherche interagissent pour développer un questionnement argumenté, fondé sur un cadre théorique, et prenant en compte les questions initiales des praticiens. Un autre point de vigilance concerne la spécification des données à collecter.

3 La plateforme co.LAB

Cet article décrit une solution technique, la plateforme co.LAB, destinée à l'accompagnement de la conduite d'un projet de ROC. Pour ce faire nous nous appuyons sur l'examen des fonctionnalités de la plateforme au regard des principes définis par Mandran et al. [4]. Nous nous appuyons également sur les résultats de l'expérimentation de la plateforme. Cette expérimentation a en particulier permis d'évaluer son utilisabilité.

Le modèle présenté plus haut (fig. 1) a orienté les choix de conception de la plateforme. Ce travail de conception s'appuie également sur l'expérience acquise dans le cadre de la conception et l'expérimentation de deux jeux. L'évaluation par inspection de la plateforme s'appuie sur l'état de l'art. Les développeurs de la plateforme ont été invités à indiquer comment les fonctionnalités de la plateforme permettent de répondre à ces différents points. La plateforme co.LAB a également fait l'objet d'une expérimentation par 15 étudiants de Master. Ces étudiants suivent, un enseignement sur les méthodes de recherche. Ils ont conduit un projet consistant à formaliser un processus de type ROC à l'aide de la plateforme. Au terme des projets, les étudiants ont répondu à un questionnaire. Une première partie était destinée à évaluer l'utilisabilité de la plateforme à l'aide de la version française de l'échelle de mesure *System Usability Scal*e (F-SUS) [11]. La seconde partie portait sur l'évaluation de fonctionnalités spécifiques et leurs attentes en termes de nouvelles fonctionnalités. Le questionnaire comprenait des questions fermées, une échelle de Likert en cinq points (1-Pas du tout d'accord à 5-Tout à fait d'accord), ainsi que des questions ouvertes (*e.g.* De quelles fonctionnalités supplémentaires auriez-vous besoin ?). La passation du questionnaire a été immédiatement suivie d'un *focus group* permettant d'éclairer les avis exprimés par les étudiants.

La plateforme co.LAB se présente sous la forme de « cartes ». Chaque carte comprend une zone éditable de type traitement de texte permettant d'insérer des images et des liens. Elle comprend également un menu éditable (« Documentation ») qui peut

renvoyer vers d'autres cartes ou des ressources externes. Un processus de conduite de la recherche sur les jeux numériques à visée éducative (fig. 2), ou « Projet », comprend des cartes distinctes pour les différentes étapes de ce processus (*e.g.* Problématiser, Concevoir, Produire des données, etc.).

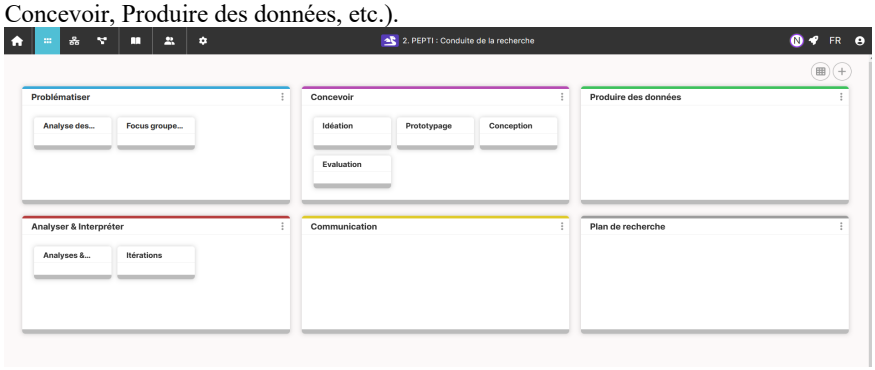


Fig. 2. Copie d'écran de la plateforme co.LAB

Chacune de ces cartes peut contenir des cartes enfants (*e.g.* Idéation, Prototypage, etc. pour la carte « Concevoir »). Les cartes peuvent être visualisées selon une vue hiérarchique ou séquentielle. Chaque carte peut être enregistrée comme carte modèle et ainsi être utilisée dans un autre projet. Une jauge permet de visualiser l'avancée du processus. Chaque utilisateur est inscrit comme invité, membre, responsable ou propriétaire. Selon les choix effectués, ses droits peuvent varier en termes d'édition (lecteur/utilisateur ou éditeur des cartes du projet). D'autres fonctionnalités permettent la sauvegarde des informations ou la consultation de la documentation.

Chaque carte comprend un résumé de l'étape du processus, les noms des membres de l'équipe impliqués dans cette étape et ce qui est attendu d'eux. Elle comprend également des liens vers des instruments nécessaires pour conduire ce travail (*e.g.* des guides) et des liens vers des livrables produits lors des étapes antérieures du processus, des liens vers la littérature relative à cette étape du processus et une zone éditable qui renvoie vers le livrable de cette étape du processus (*i.e.* de la carte).

4 Résultats et discussion

Nous indiquons ici les résultats de l'évaluation par inspection concernant la prise en compte des points de vigilance relatifs à la conduite d'une ROC. Le premier point concerne l'appui apporté par la plateforme pour que l'équipe puisse **prendre connaissance des méthodes employées et les adapter à ses propres pratiques**. C'est selon nous le point fort de la plateforme. Il est pris en compte par le fait que la plateforme permet de visualiser les activités sous la forme d'un réseau de cartes et que ces cartes sont éditables et donc adaptables pour décrire des activités spécifiques. Par ailleurs, les priorités des différentes activités peuvent être symbolisées avec un code couleur et, lorsque l'équipe travaille sur cette étape du processus, une jauge permet d'afficher le degré d'avancement. De plus, la plateforme permet d'héberger les instruments (*e.g.* guides) nécessaires. La nécessité de permettre aux experts de

communiquer les difficultés qu'ils rencontrent au reste de l'équipe n'est pas prise en compte par une fonctionnalité spécifique. Le second point concerne la **possibilité de partager des savoirs d'expertise au sein de l'équipe**. La plateforme permet ce partage car elle offre la possibilité aux experts de créer du contenu à cet effet. Par ailleurs, chaque carte permet de documenter, d'archiver et de rendre visibles les résultats des travaux effectués afin qu'ils puissent être partagés [8]. La **spécification des rôles et des compétences des acteurs du projet** fait l'objet d'une fonctionnalité spécifique. Il est possible de créer autant de rôles que nécessaire, puis d'affecter ces rôles à chaque membre à l'issue d'une négociation préalable. Les questions relatives à la **planification et à la répartition des tâches** sont prises en compte sous la forme d'une matrice RACI indiquant, pour chaque carte, qui est Responsable, Approuve, est Consulté ou Informé. L'analyse du contexte et **l'intégration d'une réflexion éthique** ne sont pas pris en compte par des fonctionnalités spécifiques, mais sont réalisables avec des fonctionnalités génériques permettant la réalisation de cartes formalisant ces aspects du processus de conduite de la recherche ou les intégrant sous la forme de recommandations dans le menu « Documentation ». Il ressort de cette analyse que la spécification, la formalisation et l'implémentation des outils et méthodes qui seront effectivement utilisés sont à la charge des utilisateurs.

Les réponses au questionnaire F-SUS permettent de calculer un score de 51/100 ce qui, au regard de l'échelle des objectifs proposée par Bangor et al. [12] la place dans la catégorie « OK », c'est-à-dire que la plateforme n'a pas suscité l'engouement de ses utilisateurs, mais elle n'a pas non plus fait l'objet d'un rejet. L'examen des travaux réalisés par les étudiants montre qu'elle leur a permis de formaliser des étapes cohérentes avec les caractéristiques et objectifs d'une ROC. Les autres réponses au questionnaire et les informations recueillies lors du *focus group* indiquent que les fonctionnalités plébiscitées sont la copie des cartes et la possibilité d'insérer des images. D'autres fonctionnalités, pas encore implémentées, ont fait l'objet de demandes très largement partagées. Il s'agit de fonctionnalités liées à l'éditeur collaboratif de texte: possibilité d'annuler des dernières modifications, stabilité de l'éditeur, mise en forme des textes et, dans une moindre mesure, système de commentaires sur les cartes. Aucune des fonctionnalités proposées dans le questionnaire n'a fait l'objet d'un rejet mais certaines d'entre elles ont reçu un intérêt limité. Il s'agit en particulier de la possibilité de recevoir des notifications automatiques, de disposer d'une vue du projet en mode liste et de la possibilité de créer des tableaux dans l'éditeur de texte.

5 Conclusion

La contribution principale de notre travail est la plateforme co.LAB et une première évaluation de son usage. Cette étude préliminaire montre que son intérêt réside dans la possibilité de définir un formalisme générique pour décrire l'ensemble des étapes d'une ROC. Nous faisons l'hypothèse que cette formalisation facilite l'appropriation du processus par l'ensemble de l'équipe. La généricité de la plateforme et de son utilisation dans le cas spécifique d'une ROC découle des fonctionnalités permettant de développer des « cartes » éditables, adaptables et partageables. Ainsi, chaque carte permet de mettre à disposition de l'équipe les méthodes et instruments nécessaires pour la conduite d'une étape de la recherche. Les points de vigilance à prendre en compte pour

conduire ce type de projet sont globalement pris en compte. Néanmoins, l'absence de fonctionnalités relatives à l'édition de texte impacte négativement son utilisabilité.

Conçue initialement pour accompagner des projets de ROC sur des jeux destinés à des usages éducatifs, il ressort que, en raison de son caractère générique, la plateforme co.LAB semble adaptée à la conduite de n'importe quel projet collaboratif de recherche-développement sur un EIAH. La possibilité d'archiver et de partager ces méthodes et instruments permettant d'alléger le travail préalable de leur spécification est probablement un atout important. Nos travaux se poursuivent pour éprouver les fondements de la plateforme en la confrontant à l'avis d'experts du domaine.

Remerciements : Ce projet a reçu le soutien du Fonds National Suisse de la Recherche Scientifique dans la cadre du Plan National de Recherche 77 et le concours des participants aux projets PRITS et TSADK, ainsi que les étudiants du Master 2 MALTT.

References

1. Sanchez, E., Monod-Ansaldi, R.: Recherche collaborative orientée par la conception. Un paradigme méthodologique pour prendre en compte la complexité des situations d'enseignement-apprentissage. Education & Didactique 9, 73-94 (2015).
2. Sanchez, E.: Enseigner et former avec le jeu. ESF Editions, Paris (2023).
3. Branch, R.: Instructional design: the ADDIE approach. Springer Science & Business Media (2009).
4. Mandran, N., Prior, E., Vermeulen, M., Sanchez, E.: Collaboration pitfalls of digital learning game multi-authors design. Education and Information Technologies (soumis).
5. Klerks, G., Slingerland, G., Kalinauskaite, I., Hansen, N., Schouten, B.: When Reality Kicks In: Exploring the Influence of Local Context on Community-Based Design. Sustainability, 14, 4107, 1-23 (2022).
6. Ke, F., Shute, V., Clark, K.M., Erlebacher, G.: Interdisciplinary design of game-based learning platforms. Springer, Cham, Germany (2019).
7. Jaccard, D., Suppan, L., Sanchez, E., Huguenin, A., Laurent, M.: The co.LAB generic framework for collaborative design of serious games: development study. JMIR Serious Games 9, (2021).
8. Prior, E.: Partage des savoirs dans une réunion de co-conception de jeux épistémiques numériques en recherche orientée par la conception. In: Bonnat, C., Venant, R. (eds.) Rencontres Jeunes Chercheurs en EIAH 2022, pp. 14-21, Lille, France (2022).
9. Baradaran Rahimi, F., Kim, B.: Playce-making: transformation of space in a participatory game design project within a Canadian junior high school. Learning. Media and Technology (2022).
10. Burton-Jeangros, C. (ed.): L'éthique (en) pratique: la recherche en sciences sociales, Vol. 34. Sociograph, Université de Genève, Genève (2017).
11. Gronier, G.B., A: Psychometric evaluation of the F-SUS: creation and validation of the French version of the system usability scale. International Journal of Human–Computer Interaction 37, 1571-1582 (2021).
12. Bangor, A., Kortum, P., Miller, J.: Determining what individual SUS scores mean: Adding an adjective rating scale. . Journal of usability studies 4, 114-123 (2009).

EIAH2023 : 11ème Conférence sur les Environnements Informatiques pour l'Apprentissage Humain
La conférence pluridisciplinaire francophone sur la conception et l'analyse des environnements numériques pour l'éducation et la formation
12-16 juin 2023 Brest (France)

Session de communications 2.1
Conception et outils pour la conception - 2ième partie

Un framework de conception pour des générateurs d'activités de jeu variées et adaptées

Bérénice Lemoine[0000−0002−7608−3223], Pierre Laforcade[0000−0001−8498−2731], and Sébastien George[0000−0003−0812−0712]

LIUM Laboratoire d'Informatique de l'Université du Mans, Laval, France
{berenice.lemoine, pierre.laforcade, sebastien.george}@univ-lemans.fr

Résumé. Cet article concerne l'aide à la conception de générateurs d'activités dans les jeux sérieux. Le champ d'étude est celui de l'entrainement aux connaissances déclaratives. La proposition est un framework de conception, orienté modèles et méta-modèles, extensible pour l'entrainement à des faits spécifiques. Le générateur produit permet de générer des activités d'entrainement adaptées et variées sous forme de niveaux de donjon pour un jeu de type Roguelite. L'article présente et évalue le framework en se focalisant sur la dimension éducative de l'entrainement.

Mots-clés : Adaptation · Génération · Jeux sérieux · Conception.

Abstract. This article deals with the design of activity generators in serious games. The field of study is declarative knowledge training. The proposal is a design framework oriented towards models and extensible metamodels for the training of specific facts. The built generator allows to generate adapted and varied training activities as dungeon levels for a Roguelite game. The article presents and evaluates the framework with a focus on the educational dimension of the training.

Keywords: Adaptation · Generation · Serious Game · Design.

1 Introduction

La répétition est une tâche nécessaire pour permettre la mémorisation à court et long terme des connaissances déclaratives [15]. Or, les jeux sérieux proposant des activités répétitives ou redondantes peuvent vite ennuyer les joueurs, d'autant plus si le challenge proposé n'est pas équilibré avec leurs compétences [24]. Ainsi, des jeux sérieux visant l'entrainement à des connaissances déclaratives doivent proposer une grande variété d'activités adaptées à l'apprenant. Pourtant, il y a très peu de solutions pour faciliter la *génération* d'activités d'apprentissage ludiques à la fois adaptées et variées.

Nos travaux concernent la recherche en ingénierie pour l'aide à la conception et au développement de générateurs d'activités variées et adaptées. Les générateurs visés sont des composants logiciels de jeux sérieux dédiés à l'entrainement pour des connaissances déclaratives. Il s'agit d'une recherche exploratoire

visant à mieux caractériser ces générateurs (l'objet de recherche) et à proposer des modèles, outils et techniques permettant de faciliter leur conception (au sens informatique) et leur développement.

Cet article a pour objectif de présenter un framework de conception qui permettra d'assister la conception et l'implémentation de ces générateurs. Cet article se concentre sur la prise en compte du point de vue de l'enseignant, de la structuration de l'entrainement, et de la progression de chaque élève. La génération prendra en compte à terme les aspects ludiques (i.e., variété de *gameplays*, préférences de jeu des apprenants).

La Section 2 présente le contexte de recherche et positionne les travaux par rapport à la littérature. La Section 3 décrit le framework de conception proposé et ses composants. La Section 4 présente une évaluation du framework sur la dimension éducative. Enfin, la Section 5 conclut et présente les perspectives.

2 Contexte et positionnement

2.1 Contexte initial : le projet AdapTABLES

Le projet AdapTABLES s'intéresse à l'entraînement aux tables de multiplication. Les tables sont considérées déjà *expérimentées* et *comprises*, l'objectif est ici de les *stabiliser* [9]. La littérature en psychologie cognitive a montré que le processus de récupération de concepts ou de faits par le test augmente leur acquisition à long terme [3,15]. Le *Retrieval Practice* est une forme d'entrainement par le test consistant à se souvenir de manière répétée de ce que l'on a appris (e.g., par l'utilisation de *flashcards*, de quiz, etc.). L'entraînement proposé ici est une forme de *Retrieval Practice* consistant à fournir à l'élève-joueur diverses formes de questions de manière répétée et variée.

Le projet vise alors à concevoir et à développer un jeu sérieux dédié à cet entrainement. La répétition d'activités similaires pouvant provoquer l'ennui [24], le jeu sérieux devra proposer à chaque élève une grande variété d'activités adaptées.

2.2 Enjeux scientifiques

L'objet de recherche se porte sur l'un des composants du jeu d'entraînement : le générateur d'activités. Celui-ci doit fournir des activités variées et adaptées à l'apprenant-joueur sur les dimensions éducative et ludique. La conception d'un tel générateur est complexe et ne peut pas être réduite à un problème d'ingénierie informatique. Cela pose de nombreux problèmes de spécification et de mise en œuvre (Quels éléments doivent être explicités et comment ? Comment peuvent-ils être exploités pour diriger la mise en œuvre informatique de la génération ?) qui impliquent les points de vue de l'équipe pluridisciplinaire de développement du jeu sérieux d'entraînement, dont les experts des connaissances déclaratives visées et les experts du jeu vidéo.

Adaptations visées. L'adaptation est souvent caractérisée par 3 concepts : la source (i.e., à quoi adaptons-nous ?), la cible (i.e., qu'est-ce qui est adapté ?) et les chemins (i.e., comment est adaptée la cible à la source ?) [26]. Dans ces travaux, nous ciblons l'*individualisation* des activités, définie par Grant et Basye comme un moyen de répondre aux compétences ou aux capacités spécifiques des élèves, y compris les besoins spéciaux, en fournissant des progressions d'apprentissage adaptées à ses besoins [12]. L'individualisation a lieu lors de la génération d'une activité. Au sens de Plass et Pawar [21], elle concerne, dans nos travaux, deux variables : une variable cognitive (i.e., niveau de l'apprenant et sa progression) et une variable motivationnelle (i.e., préférences du joueur). La génération va donc considérer de nombreuses informations (e.g., structuration de l'entrainement, les caractéristiques du jeu, le profil de l'apprenant, le profil du joueur, etc.) pour adapter les activités : 1) au niveau de connaissance et à la progression de l'apprenant, mais également 2) à ses préférences de jeu. D'autre part, les activités sont également adaptées à la stratégie d'entrainement des enseignants puisque la structuration (i.e., définie préalablement à l'utilisation) représente ses choix en termes de faits questionnés, de difficulté, de progression dans l'entrainement.

Genre de jeu. Nous avons étudié différents *genres* de jeux afin d'identifier ceux capables de maintenir l'engagement des joueurs tout en proposant des gameplays répétitifs, mais variés. Le genre de jeu Rogue*lite* répond à ces besoins. Il se caractérise principalement par la génération procédurale de donjons au contenu pseudo-aléatoire, la mort permanente (chaque mort de l'avatar impose au joueur de commencer une nouvelle partie), et la détention limitée d'éléments de jeu déblocables (e.g., personnages, objets, *powerups*, ...). En conséquence, les activités générées sont des niveaux de donjons (i.e., salles interconnectés) où l'apprentissage a lieu dans certaines salles (i.e., salle avec question). L'adaptation éducative va donc porter sur les faits rencontrés (i.e., types de faits, manière de les questionner), leur nombre, leur ordre en fonction du niveau de l'apprenant, de ses précédents résultats et de la structuration faite par l'enseignant. L'adaptation des préférences va porter sur l'activation/désactivation d'éléments de jeu (i.e., équipements débloquant différents gameplays permettant de répondre aux faits questionnés en orientant ou déplaçant des objets par exemple).

2.3 État de l'art

La conception de jeux et de jeux sérieux a fait l'objet de nombreuses méthodes, *framework* et approches [1,7]. Cependant, beaucoup d'entre eux sont essentiellement orientés vers l'analyse de jeux existants ou bien servent de guide pour réaliser une conception générale du jeu à développer [13]. De plus, aucun de ces *frameworks* ne repose sur la génération d'activités adaptées à la fois à la dimension éducative et à la dimension ludique.

Adaptation dans les jeux sérieux. Plusieurs travaux proposent des méthodes pour adapter la ludification de plateformes et contenus éducatifs [19,8]. Cependant, la conception de jeux sérieux et la ludification présentent différents chal-

lenges. La ludification implique d'ajouter une couche d'éléments de jeu à un contenu éducatif déjà existant/structuré. En revanche, la conception de jeux sérieux nécessite que la structuration du contenu éducatif et la structuration du jeu soient pensées conjointement [22]. La littérature présente des travaux cherchant à concevoir ou à soutenir la conception de contenus de jeux adaptés. Natkin et al. [20] recommandent des quêtes en fonction d'un modèle utilisateur. À partir du modèle, un ensemble de quêtes est proposé au joueur, en fonction de ses choix et des traces, le modèle est mis à jour pour permettre des propositions raffinées par la suite. Marne et al. [18] proposent un outil auteur soutenant les enseignants dans la conception de scénarios de jeux non linéaire adaptés. Les apprenants seront confrontés à différents scénarios en fonction des réponses qu'ils fournissent. Carron et Marty [6] proposent une approche pour la conception de modèle usager adapté pour des jeux sérieux. Bontchev et al. [2] proposent un framework composé d'un modèle d'apprenant-joueur pour la personnalisation de jeux sérieux de type labyrinthe découpé en 3 axes : caractéristiques du joueur, caractéristiques de l'apprenant et caractéristiques de l'utilisateur. Ces travaux sont majoritairement fondés sur l'utilisation de modèles et des résultats d'apprenants pour adapter. Cependant, tous modèles et toutes formes d'adaptation dépendent du contexte d'application et sont difficilement applicables à d'autres contextes.

Génération de contenus en EIAH. La génération consiste à proposer des contenus par construction. Peu de travaux traitent cette problématique en EIAH. Certains travaux abordent ce problème hors du contexte des jeux sérieux [5,10]. En EIAH, Callies [4] propose une architecture adaptative consistant à générer des plans pédagogiques adaptés (i.e., sous-partie d'un scénario) ainsi qu'à adapter les comportements des PNJ (Personnage Non Joueur) en fonction des actions des joueur-apprenants dans des jeux de type simulation. Sebaha et Hussaan [23] proposent un modèle générique qui génère un scénario de jeux adapté à l'apprenant en fonction de son profil et des modèles de représentation des ressources et des concepts définis. D'autre part, Laforcade et Laghouaouta [16] proposent une approche permettant d'appréhender la spécification de générateurs de séquences d'activités (i.e., scénario de jeu) adaptées à l'apprenant. Cette approche s'appuie sur trois perspectives (points de vue incrémentaux sur les éléments à générer) et trois dimensions (les éléments à générer, les éléments décrivant le contexte de la génération, les éléments décrivant le jeu d'apprentissage). Il s'agit d'une approche de spécification très généraliste dont les méta-modèles et modèles sont produits durant l'application de l'approche. En dehors des EIAH, certains travaux abordent la génération de contenus adaptée dans le cadre des jeux, comme Dormans et Bakkes [11] qui proposent un *framework* pour la conception de niveaux de jeu adaptés, de type action-aventure, fondée sur l'utilisation de *grammaires génératives*. Des niveaux à choix conditionnels sont créés dynamiquement pour chaque type de joueurs afin de personnaliser l'expérience de jeu. Bien qu'intéressant, ces travaux ne sont pas directement applicables à notre contexte. Certains sont fondés sur un principe de recommandation (i.e., différent de la génération), d'autres visent la conception de jeux sérieux mais pas leur implémentation, d'autres s'intéressent à la génération sur une seule dimension.

3 Framework de conception et d'implémentation

3.1 Présentation générale & hypothèses

Le *framework* proposé est une infrastructure conceptuelle et logicielle composée d'un ensemble de modèles et outils pour formaliser et guider l'implémentation de générateurs d'activités variées et adaptées. Il s'agit d'une contribution de recherche en ingénierie en EIAH [25] participant à explorer et à orienter des solutions concernant la génération d'activités adaptées. Notre proposition devra être complétée et ajustée pour être directement exploitable par les équipes pluridisciplinaires impliquées dans le développement de jeux d'apprentissage.

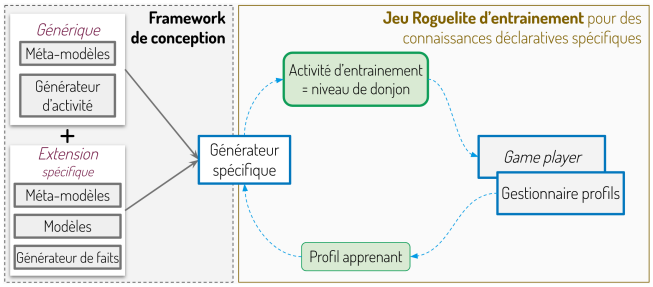


Fig. 1: Positionnement général du *framework* de conception

Comme illustré en Figure 1, le *framework* permet de spécifier et de mettre en œuvre un générateur d'activités d'entrainement pour un contexte de connaissances déclaratives spécifique. Ce générateur peut être considéré comme un composant logiciel du jeu *Roguelite* d'entrainement. Son rôle est de générer à chaque demande une nouvelle activité d'entrainement, concrètement un niveau de donjon, qui sera ensuite interprétée par le jeu. Le générateur nécessite en entrée les informations contextuelles de la demande : profil de l'élève (i.e., sa progression, ses résultats dans son parcours d'entrainement) et profil du joueur (e.g., niveau de donjon atteint, préférences de *gameplays* qu'il/elle déclare).

La conception d'un tel *framework* repose sur plusieurs hypothèses :

H1) La conception de générateurs automatiques (fonctionnels sans intervention humaine) d'activités d'entrainement de type *Roguelite* adaptés à l'apprenant sur deux dimensions (i.e., ludique et éducative) est réalisable ;

H2) Prendre en compte des connaissances déclaratives spécifiques est possible ;

H3) Les générateurs produits peuvent fournir des activités adaptées du point de vue de l'enseignant (e.g., objectifs, ordre de réalisation, tâches à travailler) ;

H4) Les générateurs produits peuvent fournir des activités adaptées au profil de l'apprenant (i.e., ses connaissances et sa progression) ;

H5) Les générateurs produits peuvent fournir des activités adaptées au profil du joueur (e.g., préférences de jeu sélectionnées dans le jeu par le joueur) ;

H6) Les générateurs produits peuvent fournir des activités variées (i.e., pour deux profils identiques, les activités doivent différer).

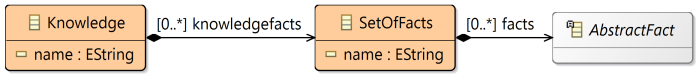
3.2 Méthode de conception

L'élaboration du *framework* est basée sur une méthode de recherche exploratoire centrée sur la conception itérative d'un cas d'étude initial, avec l'implication d'un groupe d'usagers et la participation d'experts. Il s'agit donc d'une méthode inductive où les résultats obtenus (en rapport avec le projet AdapTABLES) sont généralisés et re-évalués dans le contexte des tables de multiplication, mais également sur un second (les repères d'histoire-géographie du Diplôme National du Brevet). La continuité d'une étude exploratoire [17] menée à l'aide d'un groupe d'experts en mathématiques (i.e., didacticiens et enseignants) a permis de définir une structuration du savoir sous forme de *parcours d'entrainement*. Cette **structuration vise à expliciter** la vision de l'enseignant sur **la progression de l'entrainement** (i.e., adaptation), en termes de connaissances à travailler et de paramètres à varier, **pour un apprenant ou un groupe d'apprenants**.

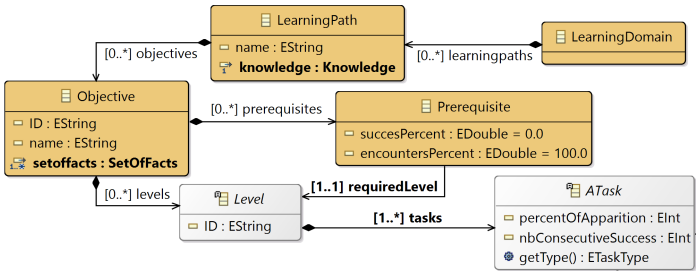
3.3 Composants du Framework

Une manière de répondre à l'hypothèse H2 (cf. Section 3.1) consiste à proposer un *framework* décomposé en une partie générique, permettant la spécification des éléments nécessaires à la génération, et une partie spécifique, représentant une extension de la partie générique et permettant de spécifier les éléments précis du domaine d'application (cf. Figure 1). La partie générique contient également un générateur déjà implémenté capable d'interpréter et de manipuler les modèles conformes aux méta-modèles génériques et leurs extensions. La génération des faits étant dépendante du domaine, elle devra faire l'objet d'un développement dédié selon des règles spécifiques afin d'être interopérable avec le reste du générateur. Notre approche vise à utiliser le cadre théorique et pratique de l'Ingénierie Dirigée par les Modèles (IDM) [14]. Le *framework* proposé est fondé sur un ensemble de méta-modèles extensibles interconnectés, ainsi qu'un ensemble de règles à suivre pour créer des extensions. Les méta-modèles sont spécifiés au format EMF et l'algorithme de génération est codé en JAVA.

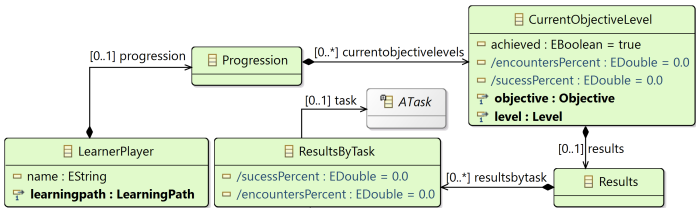
Méta-modèles "Génériques". Le *framework* est composé de plusieurs méta-modèles, les quatre principaux sont présentés dans cet article (cf. Figure 2a). Le méta-modèle des connaissances (cf. Figure 2a) permet de représenter les connaissances/faits à travailler sous forme d'ensemble de faits (*SetOfFacts*). Le méta-modèle du domaine (*LearningDomain*, cf. Figure 2b) permet de représenter la structuration de l'entrainement (*LearningPath*) en objectifs, en niveaux, en prérequis et en tâches (*ATask*). Un objectif (*Objective*) est disponible lorsque ses prérequis (*Prerequisite*) sont respectés en termes de pourcentage de succès (*successPercent*, réussite de l'apprenant calculé à partir de ses résultats) et de pourcentage de faits rencontrés (*encountersPercent*, nombre de faits minimum, défini par les enseignants, qui doivent être travaillés). Le méta-modèle de l'apprenant (*LearnerPlayer*, cf. Figure 2c) permet de représenter la progression de l'apprenant par rapport aux couples objectif/niveau (*CurrentObjectiveLevel*) d'un parcours d'entrainement. Enfin, le méta-modèle d'activité



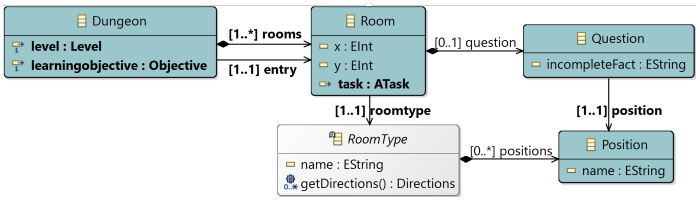
(a) MM Connaissances



(b) MM Structuration du savoir



(c) MM Apprenant



(d) MM Activité/Donjon

Fig. 2: Métamodèles génériques du *Framework* (les méta-attributs en gras indiquent une référence vers un méta-élément d'un autre méta-modèle).

(cf. Figure 2d) représente la structure que doit avoir une activité générée. En l'occurrence, l'activité est un donjon (*Dungeon*), c'est-à-dire un ensemble de salles (*Room*) pouvant avoir une question ou non, ayant un type (*RoomType*, cet attribut correspond aux paramètres d'une salle, e.g., forme). Ces méta-modèles représentent l'ensemble des informations nécessaires à la génération d'une activité d'entrainement. Dans le but d'instancier ces méta-modèles, une extension doit être apportée pour spécifier les informations du domaine visé.

Règles d'Extension. Les méta-modèles présentés (cf. Figure 2) présentent plusieurs points d'extension principalement représentés par des classes abstraites.

Du point de vue éducatif, deux méta-modèles sont visés, les méta-modèles des connaissances et le méta-modèle du domaine. Les faits peuvent prendre différentes formes en fonction des domaines. Par exemple, un fait d'une table de multiplication peut être représenté comme une classe avec trois entiers x (opérande), y (opérande) et res (résultat). En revanche, une date historique serait plutôt représentée par une classe avec une chaine de caractère (évènement) et une date ou période (entiers). En conséquence, *AbstractFact* doit être définie pour présenter la structure des faits à travailler. Les paramètres des niveaux sont entièrement dépendants du domaine visé. Par exemple, la manière de construire les tables (opérande × table ou table × opérande) ne dépendant que du domaine des mathématiques. C'est également le cas des tâches. Dans une tâche visant à compléter un fait avec un élément manquant (e.g., retrouver la date historique, ou le résultat d'une multiplication), l'élément à trouver dépend du domaine (e.g., résultat, opérande ou table pour les multiplications, ou bien évènement ou date pour l'histoire). Pour la dimension éducative, seules ces trois classes sont à étendre. Une fois l'extension conçue, une instanciation (modèle) de chaque méta-modèle, sauf celui des activités qui est instancié par le générateur, doit être créé. Dans le cas du modèle d'apprenant, le seul besoin consiste à associer chaque apprenant au bon parcours d'entrainement, la progression sera instanciée et évoluera en même temps que la progression de l'apprenant dans le jeu (i.e., par un composant du jeu gérant les profils, non présenté dans cet article).

Algorithme de Génération. La génération s'effectue en plusieurs étapes incrémentales (similaire à [23,16]). Tout d'abord, cela consiste à sélectionner un couple objectif/niveau parmi ceux éligibles. Ensuite, la structure du donjon (i.e., agencement des salles, association des salles avec les tâches) est générée en fonction des tâches à réaliser et paramétrée par l'objectif/niveau sélectionné, incluant leur pourcentage d'apparition exigée, et en fonction du nombre de salles avec question à générer. Cette génération repose sur un algorithme de *backtrack* inspiré du principe de *Grid-Based Dungeon Generator* (i.e., l'espace est divisé en cellules dans lesquelles des salles peuvent être placées). Cependant, notre approche diffère puisque nous gérons deux types de salles ($petite = 1$ cellules et $grande = 4$ cellules), créant un besoin de gestion des chevauchements de salles. Une troisième étape consiste à générer le contenu des salles en termes d'éléments de jeu (e.g., *gameplays*, objets de jeu et règles d'interaction) définis par les *gameplays*, eux-mêmes en relation avec les types de tâches associés. Les valeurs des éléments de jeu (e.g., textes à afficher, propositions portées par les objets) sont définies afin de permettre le questionnement d'un fait spécifique. La génération est donc dépendante des extensions réalisées avec le framework. Cette troisième étape très opérationnelle n'est pas présentée et évaluée dans l'article.

4 Évaluation

L'évaluation du *framework* ne repose dans cet article que sur la dimension éducative et se découpe en trois étapes : 1) l'évaluation de la prise en compte du

point de vue de l'enseignant (*H3* et, par preuve de concept, partiellement *H1* et *H2*), 2) l'évaluation de l'adaptation des activités au profil de l'apprenant (*H4*), 3) l'évaluation de la variété du contenu éducatif dans les donjons (*H6* en partie). Étant liée aux préférences de jeu, l'hypothèse *H5* n'est pas abordée.

4.1 Prise en compte du point de vue enseignant

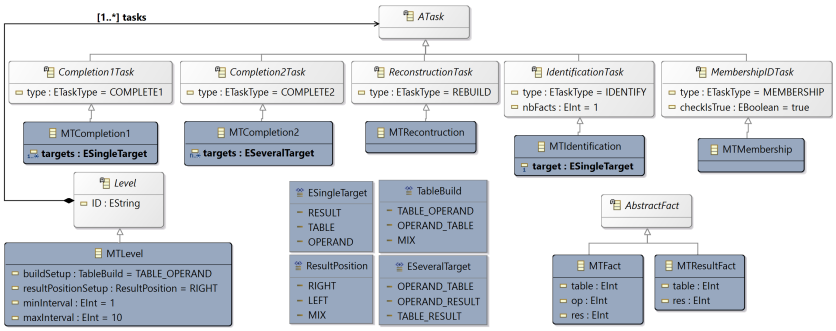


Fig. 3: Extension des Méta-modèles du Framework aux tables de multiplication

Cette évaluation s'effectue par preuve de concept en a) étendant le *framework* à un domaine spécifique (i.e., les tables de multiplication) et en b) modélisant les différentes informations (i.e., type de tâches, parcours d'entrainement, etc.). Dans ce but, nous avons travaillé avec un groupe d'usagers (enseignants des cycles 2 et 3) pour définir des tâches, types de tâches et les différents faits concernant les tables de multiplication. Cinq types de tâches ont été identifiés : **Complétion 1**, i.e., compléter un fait incomplet ayant un élément manquant (e.g., $3\times? = 15$, $15 = ?\times5$) ; **Complétion 2**, i.e., compléter un fait incomplet ayant deux éléments manquants (e.g., $3\times? = ?$ avec un ensemble de choix donnés [3, 6, 5, 15]) ; **Reconstruction**, i.e., replacer, dans l'ordre correct, tous les éléments importants d'un fait (e.g., $?\times? = ?$ avec un ensemble de choix donnés [3, 6, 5, 10, 15]) ; **Identification**, i.e., identifier l'exactitude ou l'inexactitude d'un ou de plusieurs faits (e.g., $3\times5 = 15$, vrai ou faux ?) ; **Identification d'Appartenance**, i.e., identifier les éléments qui partagent ou non une propriété donnée (e.g., [3, 5, 9, 14, 21] qui sont des résultats de la table 3 ?). Également, deux types de faits ont été identifiés : les faits classiques (i.e., trois entiers formant une multiplication) et les faits d'appartenance (i.e., deux entiers : une table et un résultat de cette table). Des paramètres spécifiques aux tables, comme l'élément cible/manquant recherché, ont été décrits. Ceci a permis d'aboutir à l'extension présentée sur la Figure 3.

Actuellement, 2 parcours d'entrainement ont été modélisés correspondant à deux enseignantes et pour deux niveaux scolaires (CE1 et CE2). Les deux

enseignantes prévoient de construire trois parcours d'entrainement correspondant à trois niveaux (i.e., approximativement faible, moyen, bon). Les parcours actuellement modélisés représentent les deux groupes de niveau "moyen". Par preuve de concept, cette modélisation permet de valider *H2* et en partie *H3*.

4.2 Adaptation à l'apprenant

L'évaluation de l'adaptation aux apprenants consiste à vérifier que les activités générées sont conformes aux prédictions déductibles d'une analyse du parcours d'entrainement associé à l'apprenant. Ceci est réalisé sous la forme de tests automatisés (JUnit). La conformité de l'activité repose sur deux points : a) la sélection du couple objectif/niveau à travailler, b) les tâches présentes en fonction du niveau sélectionné, de la progression de l'apprenant et du pourcentage d'apparition souhaité (e.g., 20% de salle travaillant T1, T1 étant une tâche du couple sélectionné, un donjon de taille 10 doit présenter 2 salles avec T1).

Dans un premier temps, nous avons créé deux profils d'apprenants (un par parcours d'entrainement). Le premier avait 4 objectifs éligibles et 2 non éligibles. Le second avait 2 objectifs éligibles pour 9 non éligibles. Pour chacun, un test unitaire a été rédigé vérifiant que sur 30 générations, tous les couples objectif/niveau éligibles étaient sélectionnés au moins une fois, mais également qu'aucun des couples objectif/niveau non éligibles n'étaient sélectionnés. Dans un second temps, pour avoir un meilleur contrôle sur les données, un parcours d'entrainement fictif à un objectif, un niveau et 6 tâches (i.e., deux tâches de type Complétion 1, mais différents paramètres) a été modélisé. Les tâches n'ont pas toutes le même pourcentage d'apparition (20%, 10%, 15%, 20%, 20%, 15% respectivement). Quatre situations critiques (i.e., cas limites) ont été vérifiées : 1) l'apprenant n'a pas commencé l'entrainement, toutes les tâches ont un taux de succès et de faits rencontrés à 0% (i.e., toutes les tâches apparaissent proportionnellement à leur pourcentage d'apparition dans le donjon) ; 2) toutes les tâches ont été démarrées, mais aucune n'est terminée (i.e., les deux pourcentages sont supérieurs à 0, toutes les tâches apparaissent proportionnellement à leur pourcentage d'apparition dans le donjon) ; 3) une seule tâche a été terminée (i.e., les deux pourcentages de cette tâche sont à 100% et la tâche n'apparait plus dans le donjon) ; 4) toutes les tâches sauf une ont été terminées (i.e., seule la tâche non terminée apparait dans le donjon). Dans le cas de la disparition d'une tâche, l'algorithme répartit son pourcentage d'apparition proportionnellement aux pourcentages d'apparition des tâches restantes (i.e., soit T1[50%], T2[20%], T3[20%], T4[10%] où T1 est achevée, alors les pourcentages deviennent T2[40%], T3[40%], T4[20%]). Les choix des couples objectif/niveau étant cohérents pour chaque profil d'apprenant, l'hypothèse *H4* est vérifiée. De plus, la bonne répartition des tâches dans les donjons vient compléter la validation de *H3*.

4.3 Variété des activités

La variété des donjons dépend à la fois du contenu éducatif et du contenu de jeu. Cet article ne s'intéresse qu'à la variété du point de vue éducatif (i.e., à

l'exception de la forme, agencement des salles, du donjon). L'évaluation sur la dimension éducative repose principalement sur la répartition des tâches dans le donjon et des faits travaillés (i.e., cette seconde partie n'est pas abordée). Grâce au projet AdapTABLES, nous disposons actuellement d'un *game player* (prototype) des activités générées. Ce *player* permet de visualiser les donjons sans avoir à les parcourir (i.e., carte visuelle où chaque salle est colorée en fonction de la tâche associée). En conséquence, l'évaluation proposée consiste à générer puis comparer visuellement 5 donjons pour chaque profil créé précédemment (i.e., les 4 profils fictifs et les 2 profils fondés sur les parcours des enseignantes). Tous les donjons générés pour un même profil se sont bien avérés différents, impliquant ainsi la validation partielle de *H6* (i.e., dimension éducative).

5 Conclusion & Perspectives

Cet article présente un *framework* de conception de générateur d'activités variées et adaptées pour l'entrainement à des connaissances déclaratives. Le *framework* a pour cible des activités pour des jeux de type *Roguelite*. L'adaptation prendra en compte à terme les dimensions éducatives et ludiques. L'objet de cet article est de présenter et d'évaluer dans un premier temps ce *framework* selon le point de vue éducatif de l'adaptation. Cette évaluation a été présentée dans le contexte de l'entrainement aux tables de multiplication et a porté sur les deux premiers niveaux de la génération (objectif et structuration).

L'élaboration du *framework* se poursuit actuellement. Le troisième niveau de génération (caractéristiques des salles) est en cours de finalisation. Il implique de prendre en compte 1) une extension spécifique au domaine visé pour générer des faits *questionnables* en fonction des tâches spécifiques paramétrées, et 2) la mise en œuvre des tâches en différents *gameplays* de réalisation. Ces *gameplays* feront alors l'objet principal des préférences ludiques pour l'apprenant-joueur. À l'avenir, nous souhaitons également appliquer ce *framework* à un second domaine (i.e., repères d'histoire géographie du DNB).

References

1. Amory, A.: Game object model version II: a theoretical framework for educational game development. Education Tech Research Dev **55**(1), 51–77 (Jan 2007)
2. Bontchev, B.P., Terzieva, V., Paunova-Hubenova, E.: Personalization of serious games for learning. ITSE **18**(1), 50–68 (May 2021)
3. Brame, C.J., Biel, R.: Test-Enhanced Learning: The Potential for Testing to Promote Greater Learning in Undergraduate Science Courses. LSE **14**(2) (Jun 2015)
4. Callies, S.: Architecture de génération automatique de scénarios pédagogiques de jeux sérieux éducatifs. Ph.D. thesis, Université du Québec à Montréal (2016)
5. Carpentier, K., Lourdeaux, D.: Generation of Learning Situations According to the Learner's Profile Within a Virtual Environment. In: Agents and Artificial Intelligence, vol. 449, pp. 245–260. Berlin, Heidelberg (2014)
6. Carron, T., Marty, J.C.: Améliorer la motivation dans les environnements d´apprentissage basés sur le jeu. Revue d'IA **25**(2), 223–251 (2011)

7. 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 (Sep 2015)
8. Codish, D., Ravid, G.: Adaptive Approach for Gamification Optimization. Proceedings - UCC 2014 pp. 609–610 (Jan 2015). https://doi.org/10.1109/UCC.2014.94
9. Dias, T.: Enseigner les mathématiques à l'école. Magnard, Paris, France (2018)
10. Diwan, C., Srinivasa, S., Ram, P.: Automatic Generation of Coherent Learning Pathways for Open Educational Resources. In: Transforming Learning with Meaningful Technologies, vol. 11722, pp. 321–334. Springer IP, Cham (2019)
11. Dormans, J., Bakkes, S.: Generating Missions and Spaces for Adaptable Play Experiences. IEEE Trans. Comput. Intell. AI Games **3**(3), 216–228 (Sep 2011)
12. Grant, P., Basye, D.: Personalized Learning: A Guide for Engaging Students with Technology. International Society for Technology in Education (2014)
13. Junior, R., Silva, F.: Redefining the MDA Framework—The Pursuit of a Game Design Ontology. Information **12**(10), 395 (Sep 2021)
14. Kent, S.: Model Driven Engineering. In: Integrated Formal Methods. pp. 286–298. Springer Berlin Heidelberg, Berlin, Heidelberg (2002)
15. Kim, J.W., Ritter, F.E., Koubek, R.J.: An integrated theory for improved skill acquisition and retention in the three stages of learning. Theoretical Issues in Ergonomics Science **14**(1), 22–37 (Jan 2013)
16. Laforcade, P., Laghouaouta, Y.: Generation of Adapted Learning Game Scenarios: A Model-Driven Engineering Approach. In: CSEDU 2018, Funchal, Madeira, Portugal, Revised Selected Papers. CCIS, vol. 1022, pp. 95–116. Springer (2018)
17. Laforcade, P., Mottier, E., Jolivet, S., Lemoine, B.: Expressing adaptations to take into account in generator-based exercisers: An exploratory study about multiplication facts. In: CSEDU 2022. Online Streaming, France (Apr 2022)
18. Marne, B., Carron, T., Labat, J.M., Marfisi-Schottman, I.: MoPPLiq: A Model for Pedagogical Adaptation of Serious Game Scenarios. In: ICALT 2013, Beijing, China, July 15-18, 2013. pp. 291–293. IEEE Computer Society (2013)
19. Monterrat, B., Yessad, A., Bouchet, F., Lavoué, E., Luengo, V.: MAGAM: A Multi-Aspect Generic Adaptation Model for Learning Environments. In: Data Driven Approaches in Digital Education, vol. 10474, pp. 139–152. Springer IP, Cham (2017)
20. Natkin, S., Yan, C., Jumpertz, S., Marquet, B.: Creating Multiplayer Ubiquitous Games using an adaptive narration model based on a user's model. In: DiGRA 2007, Tokyo, Japan, September 24-28. Digital Games Research Association (2007)
21. Plass, J.L., Pawar, S.: Toward a taxonomy of adaptivity for learning. Journal of Research on Technology in Education **52**(3), 275–300 (Jul 2020)
22. Prensky, M.: Computer Games and Learning: Digital Game-Based Learning. Handbook of Computer Game Studies (2005)
23. Sehaba, K., Hussaan, A.M.: GOALS: generator of adaptive learning scenarios. Int. J. Learn. Technol. **8**(3), 224–245 (2013)
24. Streicher, A., Smeddinck, J.D.: Personalized and Adaptive Serious Games. In: Entertainment Computing and Serious Games, vol. 9970, pp. 332–377. Cham (2016)
25. Tchounikine, P.: Précis de recherche en Ingénierie des EIAH (Jun 2009)
26. 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 (Jan 2011)

MIXAP : Un outil auteur d'activités éducatives en réalité augmentée

Mohamed Ez-zaouia[1], Iza Marfisi-Schottman[1], Cendrine Mercier[2]

[1] Le Mans Université, LIUM, Le Mans, France
[2] Nantes Université, CREN UR 2661, Nantes, France
{mohamed.ez-zaouia, iza.marfisi}@univ-lemans.fr,
cendrine.mercier@univ-nantes.fr

Résumé. La Réalité Augmentée (RA) a un grand potentiel pour l'apprentissage multisensoriel et expérientiel. Cependant, la création d'activités éducatives en RA est loin d'être une tâche facile pour les non-informaticiens. Pour examiner si une approche outil auteur pour la RA peut être bénéfique en contexte éducatif, nous avons conçu MIXAP dans une démarche participative auprès de 19 enseignants pilotes. Nous avons mené une étude avec 39 enseignants pour évaluer l'outil avec des enseignants pilotes et non-pilotes. Nous avons trouvé que cette approche peut aider les enseignants à créer des activités pédagogiques de RA. Les tailles d'effets de la facilité de prise en main, le ressenti émotionnel, et la faible charge cognitive de MIXAP étaient significativement importantes avec un transfert entre enseignants pilotes et non-pilotes. Nous discutons de nos résultats et proposons des perspectives.

Mots-clés : Outil Auteur, Réalité Augmentée, Éducation, Design Itératif, Design Participatif

Abstract. Augmented Reality (AR) has great potential for multisensory and experiential learning. However, creating educational activities in AR is far from an easy task for non-technical users. To examine whether an AR authoring approach can be beneficial in an educational context, we designed MIXAP through a participatory process with 19 pilot teachers. We conducted a study with 39 teachers to evaluate the tool with both pilot and non-pilot teachers. We found that this approach can help teachers create pedagogical AR activities. The effect sizes of ease of use, emotional experience, and low cognitive load of MIXAP were significantly important with transferability between pilot and non-pilot teachers. We discuss our results and propose perspectives.

Keywords: Authoring Tool, Augmented Reality, Education, Iterative Design, Participative Design

1 Introduction

La Réalité Augmentée (RA) est de plus en plus utilisée pour l'apprentissage [6] en offrant un environnement multimodal pour représenter un contenu pédagogique et interagir avec ce dernier de manière immersive [18]. Cela peut être un bon support à la pédagogie, car elle permet aux apprenants d'interagir avec des concepts parfois difficiles à observer (ex. : force magnétique) et d'accéder à des ressources multimodales tout en manipulant des objets réels [5]. Ces modalités peuvent favoriser l'apprentissage multisensoriel (la vue, le toucher et l'ouï) et expérientiel [19, 21] dans de nombreuses disciplines [11, 2]. La recherche préconise la RA pour offrir un apprentissage immersif qui engage les apprenants et facilite la compréhension du contenu et les phénomènes étudiés [3].

Cependant, la création d'activités pédagogiques en RA est loin d'être une tâche facile. La création de contenu en RA nécessite des connaissances et des compétences avancées dans les librairies spécialisées [3]. Les ***outils auteurs*** [13] offrent une approche qui pourrait démocratiser la RA en éducation puisqu'ils permettent à des non-informaticiens, comme des enseignants, de créer leur propres applications de RA en utilisant des interactions intuitives [25]. Par exemple, un enseignant peut créer une activité de RA en prenant une photo d'un objet, d'une affiche ou d'un livre et en y ajoutant des ressources multimodales sous forme d'augmentations (texte, audio, vidéo, images ou modèles 3D). Ces ressources apparaissent quand les apprenants placent la tablette ou le smartphone au-dessus de l'objet pris en photo initialement. Bien que les outils auteurs réduisent le verrou technique, de récentes études systématiques ont soulevé plusieurs défis dans la création d'application de RA à des fins pédagogiques. En effet, les outils actuels restent complexes à utiliser pour les enseignants et la personnalisation des contenus est encore très contrainte [1, 22, 11]. En outre, des études ont révélé que les outils de création de RA existants offrent un support limité pour les contenus éducatifs [25, 6], suggérant que plus de recherches sont nécessaires. Nous avons pour objectif de répondre à la question de recherche : *Comment aider les enseignants à créer leurs propres activités pédagogiques en RA ?*

Pour examiner si et comment une approche *"outil auteur"* pour la RA peut être bénéfique en contexte éducatif, nous avons d'abord conçu Mixap (Figure 1), un prototype qui permet aux non-informaticiens de créer des activités en RA d'une manière interactive et visuelle. Nous avons conçu Mixap via un processus de design itératif et participatif avec 19 enseignants. Pour examiner notre approche avec un public plus large d'enseignants nous avons mené une étude évaluant cette approche avec 39 enseignants sur les plans d'utilisabilité, d'utilité, d'acceptabilité et du transfert. Cette étude compare un groupe d'enseignants pilotes et un 2e groupe d'enseignants non-pilote (utilisant Mixap pour la première fois). Nous avons constaté que Mixap peut aider les deux groupes d'enseignants à créer des activités de qualité en RA. Nous avons trouvé que, pour les deux groupes, la facilité d'utilisation, la facilité de prise en main et le ressenti émotionnel étaient significativement plus élevées que la valeur médiane, avec

3. Tels que Unity, Threejs, Vuforia, ARCore, etc.

des tailles d'effet fortes. De plus, nous avons trouvé que la charge cognitive de l'utilisation de Mixap était significativement inférieure à la valeur médiane.

2 État de l'art

Avantages de la RA en éducation. Plusieurs études ont été menées sur l'usage, les tendances et les avantages des applications de RA pour l'éducation [17, 1, 2, 6, 10, 11, 8]. Par exemple, Radu [17] a analysé 26 études comparant l'apprentissage avec ou sans RA et a découvert que la RA présentait plusieurs avantages pour les apprenants, tels que la motivation, la collaboration, la mémorisation et l'apprentissage de structures spatiales. Garzón et al. [8] a examiné l'impact des facteurs pédagogiques qui sont favorisés par l'usage de la RA, tels que les approches pédagogiques mises en place (ex: l'apprentissage collaboratif, l'apprentissage par projet, l'apprentissage situé, l'apprentissage multimédia), la durée de l'activité et le contexte d'utilisation (ex. : dans une salle de classe, à l'extérieur, lors d'excursions, dans des musées). Les auteurs ont notamment constaté que la RA collaborative avait l'impact le plus positif sur les apprenants. Ibáñez and Delgado-Kloos [11] a examiné la littérature de la RA dans domaines de la science, technologie, ingénierie et mathématiques et a caractérisé les applications de RA, les processus pédagogiques, les approches de recherche et les limites des travaux dans cette direction. Dans cet article, nous nous intéresserons plutôt à la production de ces applications de RA par les enseignants à travers des outils auteurs [25]. De tels outils visent à rendre la création, la modification ou l'extension d'artefacts logiciels moins techniques et plus faciles pour les non-informaticiens [13]. Bien que ces outils auteurs soient extrêmement prometteurs pour une adoption plus large de la RA dans l'éducation, elles n'ont pas été étudiées de manière approfondie.

Outils auteurs de RA. Très peu d'études ont examiné les aspects de conception pédagogique de RA. Nebeling and Speicher [16] ont classé les outils facilitant le prototypage rapide d'expériences de RA/VR en fonction de quatre catégories principales : types d'écran, interaction (utilisation de la caméra), contenu 3D et jeux 3D. Mota et al. [15] ont examiné les outils de création en fonction des paradigmes de création (autonome, plug-in) et des stratégies de déploiement (spécifique à la plate-forme, indépendante de la plate-forme). Dengel et al. [6] ont examiné 26 librairies de RA citées dans des travaux scientifiques. Ils les ont caractérisé en fonction du niveau de compétences en programmation requis (élevé, faible ou moyen), du niveau d'interactivité (statique ou dynamique), de l'accessibilité (gratuite ou commerciale), de la compatibilité des dispositifs (mobile, bureau, HMD ou Web) et de la capacité de collaboration (oui ou non). Cependant, les recherches susmentionnées se sont principalement concentrées sur les outils de création qui nécessitent un certain niveau de programmation et analysent exclusivement les outils issus de travaux de recherche, y compris les outils non éducatifs. Ez-Zaouia et al. [25] ont récemment analysé 21 outils auteurs de RA spécifiquement dédiés à l'éducation, ne nécessitant pas de programmation, issue de l'industrie et de la recherche. Les auteurs ont analysé des

outils selon une grille de quatre dimensions : (1) les modes de conception offertes (type de production de RA, sources de contenu, conception collaborative et type de plateforme), (2) les modalités de RA disponibles (suivi d'objet, augmentation d'objet, interaction et navigation), (3) l'utilisation de RA (type de dispositif, collecte de contenu, connectivité et langue) et (4) les options de gestion du contenu et des utilisateurs (partage, administration et licence). En outre, ces études ont soulevé plusieurs problèmes de conception de la RA en éducation, notamment la difficulté d'utilisation, le manque de personnalisation, le coût élevé de la technologie et l'absence de modèles holistiques et de principes de conception pour la RA [25, 6, 1, 22, 11, 16]. Bien que ces études aient fourni un aperçu sur le design des outils auteurs de la RA, il manque des travaux spécifiques aux outils auteurs de RA pour l'éducation qui impliquent des acteurs du terrain selon une méthode *Design-Based Research*.

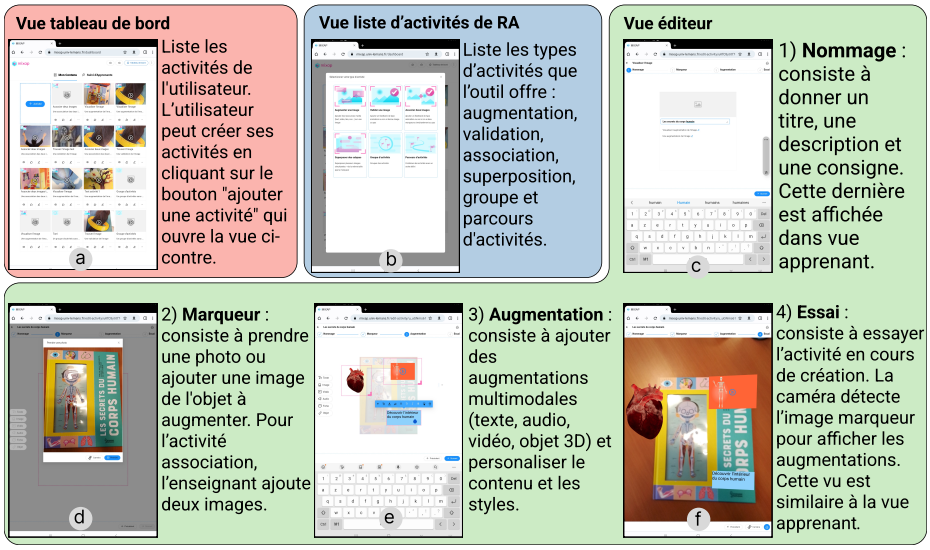


Fig. 1 – Mixap offre des workflows de création pour les six activités de RA. Chaque workflow fournit des étapes de création à l'aide d'interactions visuelles. Les augmentations de RA peuvent êtres : textes, images, audios, vidéos, objets 3D et boutons qui ouvrent des fenêtres d'informations. Les utilisateurs peuvent personnaliser le contenu et l'apparence des augmentations RA ; les tourner en 3D, redimensionner, personnaliser les styles et glisser-déposer autour d'un marqueur.

3 Processus de conception itératif centré sur l'enseignant

Nous menons nos travaux dans le cadre d'une méthode *Design-Based Research* avec un processus de conception itératif, centré sur les enseignants [23]. Nous présentons l'*exploration du domaine* en collaborant avec des experts, le *design participatif* en collaboration avec des enseignants pilotes, le *prototypage itératif* ainsi qu'une première *évaluation* de notre approche avec 39 enseignants.

3.1 Partenariat avec le réseau Canopé

Pour comprendre l'utilisation de la RA, les pratiques des enseignants et les défis liés à l'intégration de la RA dans les salles de classe au quotidien, nous avons signé un partenariat avec le réseau CANOPÉ, un organisme public qui offre des formations professionnelles aux enseignants dans toute la France. Les animateurs de CANOPE de la ville de Laval proposaient déjà des formations sur la RA et étaient en relation avec des enseignants locaux qui souhaitent utiliser cette technologie dans leur classe. Au début du projet, nous avons donc pu apprendre de leurs expériences sur les méthodes de travail qui sont les moins contraignantes pour les enseignants, les applications et les outils auteurs de RA actuellement disponibles, mais surtout les défis auxquels les enseignants font face avec ces outils. Un des défis récurrents évoqué est l'impossibilité de personnaliser le contenu. En effet, de nombreuses applications de RA, présentées comme pédagogiques, proposent du contenu prêt à l'emploi. Même si elles sont impressionnantes au premier abord, elles sont rarement adaptées aux besoins des enseignants, soit parce que le contenu n'est pas adapté aux profils des apprenants ou ne s'intègre pas de façon logique dans leurs séquences pédagogiques. Quelques outils auteurs existent tout de même, permettant aux enseignants de créer leur propre contenu [25] mais, ils ne proposent pas d'approches pédagogiques. Ces constats corroborent des résultats des revues systématiques [6, 1, 22, 11, 16]. D'autres défis sont plus prosaïques. Par exemple, la plupart des outils auteurs disponibles nécessitent une connexion internet ce qui est rédhibitoire pour plus de la moitié des enseignants français qui n'ont pas accès à un réseau WIFI stable dans leurs établissements.

3.2 Implication d'enseignants pilotes

Par le biais du réseau CANOPÉ, nous avons recruté un panel de 19 enseignants pilotes avec des profils différents puisque l'objectif est de concevoir un outil auteur qui répond aux besoins de tous les enseignants : [Sexe : (femmes = 8, hommes = 11), années d'enseignement : (min = 2, max = 40), niveau scolaire : (élémentaire = 5, moyen = 11, secondaire = 2, universitaire = 1)].

Ces enseignants pilotes ont été impliqués à travers une entrée par projet. C'est-à-dire que nous leur avons demandé d'imaginer des activités en RA qu'ils vont ensuite utiliser dans leurs classes. Les productions des enseignants et leurs réponses aux questionnaires concernant leurs profils et objectifs pédagogiques sont accessibles librement sur Mendeley [14]. L'analyse de ces productions a permis d'identifier six types d'activités pédagogiques de RA récurrentes. Le fait de partir des besoins de terrains a non seulement permis de trouver un sens pédagogique aux activités de RA mais aussi de découvrir certaines façons d'utiliser la RA qui ne sont pas faisables dans les outils auteurs existants, mais qui semblent pourtant pertinentes d'un point de vue pédagogique.

Le type d'activité que l'on retrouve le plus souvent est l' **augmentation d'image**. Elle permet de faire apparaître des ressources multimodales (texte, image, vidéo, audio, modèle 3D et fiche d'information complémentaire) sur une

image (ex. : poster, livré, fiche d'exercice). Par exemple, une enseignante de primaire souhaite ajouter des photos d'animaux et des enregistrements audio avec le vocabulaire à retenir (ex. :"Le renne a quatre sabots et deux bois") aux pages d'un livre. Un autre enseignant au lycée souhaite afficher les modèles 3D que ces élèves ont créés sur le trombinoscope de la classe. La méthode de recueil de besoins a aussi permis de révéler d'autres types d'activités plus originaux tels que la **validation d'image**. Celle-ci permet de créer des activités en autonomie, en utilisant la RA pour valider automatiquement si l'image choisie est correcte. Par exemple, un professeur d'SVT au collège veut demander aux élèves d'assembler correctement les pièces d'une carte de la pangée. Un autre veut que les élèves identifient une partie spécifique d'une machine (ex: le moteur). L'activité de type **association de deux images** permet aussi de créer des activités en autonomie puisqu'elle fait apparaître une information si les élèves ont bien trouvé deux images associées. Par exemple, un enseignant du primaire veut que les enfants s'entraînent à reconnaître la même lettre, écrite en majuscule et en minuscule. Enfin, l'activité de type **superposition d'images** permet de faire apparaître des calques transparents sur une image. Par exemple, un professeur de géologie à l'université veut que les élèves soient capables d'activer ou de désactiver des couches montrant différents types de roches sur la photo d'une montagne. D'autres enseignants souhaitent utiliser ce type d'activité pour montrer les réponses à un exercice. Pour finir, l'analyse des prototypes papier a révélé que les enseignants souhaitent regrouper les activités présentées ci-dessus en **groupe d'activité**, sans ordre prédéfini, ou sous la forme d'un **parcours d'activités** ordonnées.

3.3 Prototypage itératif et évaluations préliminaires

Nous avons conçu trois prototypes de l'outil auteur Mixap, sur une période de cinq mois. La dernière version est présentée dans la Figure 1. Nous avons décidé d'utiliser les technologies Web pour la RA pour faciliter l'adaptation aux différents dispositifs (tablettes, ordinateurs) disponibles dans les classes avec un fonctionnement en local qui ne nécessite pas d'Internet (une des contraintes techniques fortes identifiées). Ces prototypes ont été améliorés avec les enseignants pilotes (interfaces et textes descriptifs). Nous avons également testé l'application avec des élèves de maternelle qui ne savent pas encore lire pour améliorer le côté intuitif des interfaces. Nous avons également souhaité impliquer des enseignants non-pilote (NP) pour vérifier si les types d'activités identifiés par les enseignants pilotes (P) couvraient bien les besoins d'autres enseignants (section suivante).

4 Méthode: Évaluation de Mixap auprès des enseignants

Hypothèses : Nous examinons 7 hypothèses principales :
— *Utilisabilité :*
  — **H1 :** La facilité d'utilisation de Mixap sera supérieure à la valeur médiane.
  — **H2 :** La facilité de prise en main de Mixap sera supérieure à la valeur médiane.

— *Utilité :*
  — **H3 :** L'utilité de Mixap en termes de RA sera supérieure à la valeur médiane.
— *Acceptabilité :*
  — **H4 :** Le ressenti émotionnel de Mixap sera supérieur à la valeur médiane.
  — **H5 :** La charge cognitive de Mixap sera inférieure à la valeur médiane.
— *Transfert :*
  — **H6 :** Il n'y aura pas de différences entre les enseignants P et NP, au niveau d'utilisabilité, utilité, ressenti émotionnel et charge cognitive.
  — **H7 :** Il n'y aura pas de différences d'âge et de disciplines entre enseignants au niveau d'utilisabilité, utilité, ressenti émotionnel et charge cognitive.

Participants : Trente-neuf (N=39) enseignants [Age : 1 = 26-30, 1 = 31-35, 3 = 36-40, 15 = 41-45, 12 = 46-50, 7 > 50 ; disciplines : 22 = STEM, 7 = Non-STEM, 1 = Informatique, 9 = Autre]. Les enseignants sont réparti en deux groupes : le groupe P (pilote) avec 12 enseignants et le groupe NP (non-pilote) avec 27 enseignants non-familiers avec l'outil, recrutés dans le cadre d'une journée de formation professionnelle. Tous les enseignants ont volontairement participé à l'étude sans compensation et signé un accord d'utilisation des données produites pour la recherche.

Activités pédagogiques de RA : Les participants devaient créer quatre activités en RA : (1) annoter la page d'un livre avec des ressources multimodales, (2) afficher des informations pour une association de deux cartes, (3) valider un poster et (4) créer un groupe ou un parcours d'activités à partir des trois premières activités (voir ce lien).

Appareillage : Mixap été installé sur 12 de nos tablettes (Samsung SM-X200, SM-T500, 11 pouces, 2G RAM, 32G DISQUE). Les enseignants avaient également à disposition un ensemble d'accessoires pour créer des ressources pour les activités : des cartes du jeu *Dixit*, des livres et des affiches.

Procédure : Nous avons (1) présenté l'étude et les quatre activités de RA aux participants ; (2) demandé aux participants de réaliser les quatre activités ; et (3) demandé aux participants de répondre à un questionnaire (voir ce lien).

Collecte et analyse des données : Nous avons recueilli 39 réponses au questionnaire. Pour l'utilisabilité (**H1-2**), nous avons utilisé les 10 questions du SUS [Q1-10] pour évaluer le score total du SUS ainsi que la facilité l'utilisation [Q1,2,3,5,6,7,8,9] et la facilité de prise en main [Q4,10] [4]. Le SUS est un instrument unidimensionnel, toutefois c'est possible d'explorer sa bidimensionnalité [9, 12]. Nous avons formulé une [Q15] pour évaluer l'utilité (**H3**) ainsi qu'une question ouverte [Q16] qui portait sur les usages pédagogiques que les enseignants souhaitent. Pour l'acceptabilité (**H4-5**), nous avons formulé [Q12,14] pour évaluer le ressenti émotionnel et [Q11,13] pour évaluer la charge cognitive. Pour le transfert (**H6-7**), nous avons analysé la variance des réponses vis-à-vis les facteurs Groupe, Disciple et Âge. À l'exception de la Q16, les questions étaient sur une échelle de Likert à 5 points. La valeur (Likert = 3, valeur "neutre") est la médiane.

5 Résultats

Nous rapportons à la fois les p-values pour la signification statistique et les tailles d'effet (faible : < 0,3, moyenne : 0,3 - 0,5, importante : > 0,5), avec un intervalle de confiance de 95%. Nous avons utilisé un test non paramétrique de Wilcoxon pour les tests à un échantillon. Nous avons utilisé des analyses factorielles pour l'analyse de la variance des réponses de Likert [20] au niveau des facteurs Groupe, Disciple et Âge.

Vue d'ensemble : Dans l'ensemble, presque tous les participants ont terminé les activités à peu près en même temps. La Figure 2 résume les réponses au questionnaire, ce qui met en évidence les réactions positives des participants envers l'outil.

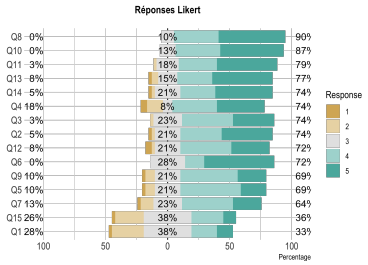


Fig. 2 – Réponses Likert.

Utilisabilité : Nous avons suivi [4, 12] et calculé les scores (de 0 à 100) pour le SUS [Q1-10], la facilité d'utilisation [Q1,2,3,5,6,7,8,9] et de prise en main [Q4,10]. Comme l'illustre Figure 3 (légende P_NP), le score moyen du SUS de tous les participants était de 73,91, ce qui est supérieur à la valeur standard du SUS de 68 points et est classé comme "bon." Un test de Wilcoxon à un échantillon a montré que les scores des enseignants étaient significativement supérieurs à 68 points, avec une taille d'effet moyenne pour le SUS total (p = 0,015, r = 0,389), une taille d'effet moyenne pour la facilité d'utilisation en main (p = 0,049, r = 0,315) et une taille d'effet importante pour la facilité de prise en main (p = 0,009, r = 0,756).

L'analyse par Groupe a montré que la différence n'était pas significative pour le SUS, avec une taille d'effet moyenne pour le groupe pilote (P) (p = 0,107, r = 0,476) et une tendance à la signification pour le groupe non-pilote (NP) (p = 0,062, r = 0,360). La différence dans la facilité d'utilisation n'était pas significative avec une taille d'effet moyenne pour les deux groupes : P (p = 0,288, r = 0,318) et NP (p = 0,104, r = 0,314). En revanche, la différence entre les groupes pour la prise en main était significative, avec une taille d'effet importante pour le groupe P (p = 0,022, r = 0,756) et une taille d'effet moyenne pour le groupe NP (p = 0,022, r = 0,443).

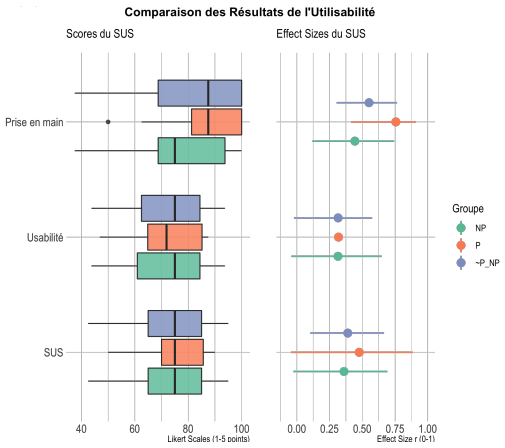


Fig. 3 – Utilisabilité : A gauche un boxplot avec la médiane, min-max et les outliers ; à droite les tailles d'effet avec l'intervalle de confiance à 95%.

Utilité : Nous avons effectué un test de Wilcoxon à un échantillon pour évaluer si l'utilité [Q14] est supérieure à la valeur médiane de 3 (Figure 5, légende P_NP). Pour les données combinées, l'utilité n'est pas significative avec une taille d'effet faible (p = 0,25, r = 0,186). L'analyse par Groupe a montré que la différence dans l'utilité était significative pour le groupe pilote (P) avec une taille d'effet importante (p = 0,043, r = 0,598) et non significative pour le groupe non-pilote (NP) (p = 0,684, r = -0,084). Certains enseignants ont élaboré sur des nouvelles fonctionnalités qu'ils souhaitent avoir dans l'outil, comme, par exemple, une activité RA capable de reconnaître plusieurs images-marqueur automatiquement et d'afficher les augmentations, l'ajout de vidéo dans des fiches d'information, ou trouver des détails dans une

grande image. D'autres ont élaboré sur des problèmes techniques tels que la précision, la stabilité ou la réactivité des augmentations. Ces éléments peuvent impacter l'utilité.

Acceptabilité : Comme illustré dans Figure 4, un test de Wilcoxon à un échantillon a montré les réponses likert étaient significativement différent de la médiane avec une taille d'effet importante pour le ressenti émotionnel [Q12,14] ($p < 0{,}001$, $r = 0{,}666$) et une taille d'effet importante (négative) pour la charge cognitive [Q11,13] ($p < 0{,}001$, $r = -0{,}827$). L'analyse par Groupe, a montré que la différence dans le ressenti émotionnel était significative avec une taille d'effet importante pour le groupe pilote (P) ($p < 0{,}0001$, $r = 0{,}814$) et non-pilote (NP) ($p < 0{,}0001$, $r = 0{,}631$). La charge cognitive était significativement inférieure à la valeur médiane de 3 avec une taille d'effet importante pour le groupe pilote (N) ($p < 0{,}0001$, $r = -0{,}898$) et non-pilote (NP) ($p < 0{,}0001$, $r = -0{,}825$).

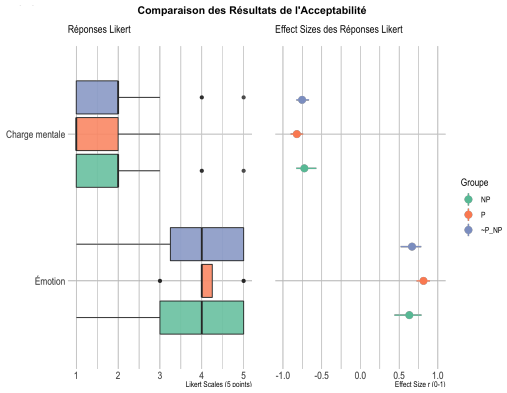


Fig. 4 – Acceptabilité : A gauche un boxplot avec la médiane, min-max et les outliers ; à droite les tailles d'effets avec l'intervalle de confiance à 95%.

Transfert : Nous avons effectué une analyse factorielle de la variance des réponses Likert via la procédure ART [20]. L'analyse a montré qu'il n'y a pas de différence significative dans les réponses entre les niveaux de la Catégorie : facilité d'utilisation, facilité de prise en main, utilité, ressenti émotion et charge cognitive (p = n.s) vis-à-vis du groupe ($p = 0{,}180$), de l'âge ($p = 0{,}889$) et de la discipline ($p = 0{,}520$).

6 Discussion

Impact de l'approche outil auteur : Les résultats présentés dans la section précédente permettent de valider un certain nombre de nos hypothèses. L'analyse des scores de SUS valide **H1**. La facilité de prise en main était très élevée, validant ainsi **H2**. L'utilisabilité de Mixap est donc raisonnable et comparable entre les enseignants pilotes et non pilotes. De plus, les deux groupes ont réussi à créer les quatre activités en à peu près le même temps donc la courbe d'apprentissage de Mixap est rapide. L'acceptabilité de l'outil est élevée puisque les participants ont montré des réactions positives (aussi observées lors des expérimentations) et leurs ressentis émotionnels étaient significativement forts, validant **H3**. Le fait d'avoir proposé des workflows et des interactions intuitives semble également avoir été utile puisque la charge cognitive était significativement faible, validant **H4**. Concernant le transfert, l'analyse factorielle de la variance n'a pas montré de différence significative entre les enseignants pilotes et non-pilotes, validant ainsi **H6** et **H7** ce qui montre que cette approche peut être bénéfique pour un public d'enseignant plus large. En revanche, l'utilité perçue était significativement élevée seulement pour le groupe pilote, ce qui valide partiellement **H5**. Dans nos prochaines itérations, nous allons donc creuser davantage les enjeux d'utilité en particulier des questions (1) d'appropriation et (2) de l'impact longitudinal de l'approche. Ce résultat sera certainement amélioré grâce aux vidéos tutorielles qui

seront associées à Mixap et qui montreront les différentes activités pédagogiques qui peuvent être créées avec.

Design participatif d'outils auteurs de RA : Un outil auteur vise à offrir une palette de fonctionnalités pour créer des applications numériques, de façon intuitive, et qui répondent aux besoins d'une large audience. Dans ce sens, un design participatif avec des enseignants nous a permis d'identifier des besoins pédagogiques que nous avons pu transposer en petites briques de base pour créer des activités en RA d'une manière simple. L'analyse du transfert a montré que cette approche est viable. En effet, même si l'outil a été conçu que sur l'analyse de besoin de 19 enseignants pilotes, l'évaluation avec 27 enseignants non-pilotes a montré qu'il n'y avait pas de différence entre les enseignants pilotes et non-pilotes, sauf au niveau de l'utilité perçue.

Cela dit, l'approche participative pose aussi certains défis. Arriver à agréger des besoins variés sans pour autant complexifier l'outil peut s'avérer être une tâche difficile. Il a été nécessaire de faire des choix dans la sélection, la validation et même le rejet de certains besoins (ex. des besoins spécifiques à un contexte d'usage ou qui ne répondent pas à des besoins communs ou risquent de complexifier l'outil pour la majorité des enseignants). Ces choix ont été discutés avec les enseignants pilotes. Certains enseignants peuvent tout de même avoir du mal à identifier leurs besoins dans l'outil —tels qu'ils les avaient exprimés, ce qui pourrait impacter négativement leur perception envers l'utilité de l'outil. Notre analyse de l'utilité supporte ce point. Certains enseignants n'ont pas identifié leurs besoins dans l'outil d'une manière explicite. Un outil auteur demande un effort d'appropriation pour adapter ses fonctionnalités afin de répondre à des besoins personnels.

Conception d'activités pédagogiques en RA : Au-delà des aspects techniques, la création d'activités en RA par les enseignants pose une certaine difficulté cognitive. Cette difficulté vient du fait que les enseignants ont peu de familiarité avec la création de contenu en RA. Ce type de productions n'est pas encore habituel, comparé, par exemple, aux productions multimédias via *Powerpoint*. Les enseignants peuvent avoir du mal à élaborer des activités pédagogiques en RA ou à adapter leurs activités traditionnelles. Même si un outil auteur enlève la barrière technique, les enseignants ont donc besoin d'accompagnement pour comprendre comment concevoir des activités pédagogiques avec ce nouvel outil.

Potentiellement, une bibliothèque d'activités pédagogiques où les enseignants peuvent trouver, créer et partager des activités de RA pourrait aider à relever les défis susmentionnés. Notre approche va dans ce sens puisque Mixap propose déjà une fonctionnalité importante qui permet de créer des copies d'activités créées par d'autres enseignants et de les modifier. Ce partage de ressources intéresse d'ailleurs beaucoup les enseignants pilotes qui ont naturellement trouver l'inspiration dans les idées de leurs collègues. Cette évolution peut être constatée en comparant leurs fiches projets initiales et les projets finaux auxquels ils aboutissent.

Limitations : La taille de notre échantillon (N=39) est relativement faible. Une analyse de la qualité des créations des enseignants et aussi des outliers manque. Un autre aspect important qui manque et qui peut améliorer l'appropriation des outils de création de RA par les enseignants est les learning analytics (par exemple, les tableaux de bord). Nous pensons que le fait de fournir aux enseignants des informations sur les expériences des apprenants, telles que l'état émotionnel [26, 24] et la progression et l'engagement [7] peut les aider a mieux approprié la RA en classes.

7 Conclusion

La RA offre un environnement intéressant pour faciliter un apprentissage multimodal, immersif et engageant. Cependant, la création d'activités en RA est loin d'être une tâche facile. Nous avons examiné une approche d'"outil auteur" pour rendre la RA moins technique et plus accessible aux enseignants (non-informaticiens). Nous avons présenté notre processus pour concevoir Mixap dans une démarche de design itératif et participatif avec des enseignants pilotes. Nous avons évalué notre approche auprès de 39 participants. Les résultats sont très encourageants pour creuser davantage cette approche, notamment l'analyse de l'usage, l'appropriation et les activités créées par les enseignants dans leurs classes. Nous espérons que ce travail fournira aux chercheurs et aux concepteurs des idées de design et d'usage d'outils auteurs et aidera à la démocratisation de la RA en éducation.

Remerciement

Nous remercions le programme étoiles montantes des Pays de la Loire, France, pour le financement du projet de recherche MIXAP. Nos remerciements spéciaux vont aux enseignants, notamment aux pilotes : Annabel Le GOFF, Camille POQUET, Damien DUMOUSSET, Delphine DESHAYES, Elisabeth PLANTE, Frederic LLANTE, Laurent HUET, Nicolas JOUDIN, Tony NEVEU, Vanessa FROC, Yannick GOURDIN, Régis MOURGUES, Morgane ACOU-LE NOAN et Adeline JAN.

References

1. Akçayır, M., and Akçayır, G.: Advantages and challenges associated with augmented reality for education: A systematic review of the literature. Educational Research Review 20, 1–11 (2017)
2. Arici, F., Yildirim, P., Caliklar, Ş., and Yilmaz, R.M.: Research trends in the use of augmented reality in science education: Content and bibliometric mapping analysis. Computers & Education 142, 103647 (2019)
3. Billinghurst, M., and Duenser, A.: Augmented Reality in the Classroom. Computer 45(7), 56–63 (2012)
4. Brooke, J., *et al.*: SUS-A quick and dirty usability scale. Usability evaluation in industry 189(194), 4–7 (1996)
5. Da Costa, J., Szilas, N., and Mueller, A.: Réalité augmentée pour l'apprentissage conceptuel en sciences: quels principes de conception pour les EIAH? Cas du dispositif DEAPE Learn en électromagnétisme. In: Actes de la 9ème Conférence sur les Environnements Informatiques pour l'Apprentissage Humain, pp. 181–192 (2019)
6. Dengel, A., Iqbal, M.Z., Grafe, S., and Mangina, E.: A Review on Augmented Reality Authoring Toolkits for Education. Frontiers in Virtual Reality 3 (2022)
7. Ez-zaouia, M., Tabard, A., and Lavoué, E.: PROGDASH: Lessons Learned from a Learning Dashboard In-the-wild. In: Proceedings of the 12th International Conference on Computer Supported Education - Volume 2: CSEDU, pp. 105–117 (2020)
8. Garzón, J., Kinshuk, Baldiris, S., Gutiérrez, J., and Pavón, J.: How do pedagogical approaches affect the impact of augmented reality on education? A meta-analysis and research synthesis. Educational Research Review 31, 100334 (2020)

9. Gronier, G., and Baudet, A.: Psychometric evaluation of the F-SUS: Creation and validation of the French version of the System Usability Scale. International Journal of Human–Computer Interaction 37(16), 1571–1582 (2021)
10. Hincapie, M., Diaz, C., Valencia, A., Contero, M., and Güemes-Castorena, D.: Educational applications of augmented reality: A bibliometric study. Computers & Electrical Engineering 93, 107289 (2021)
11. Ibáñez, M.-B., and Delgado-Kloos, C.: Augmented reality for STEM learning: A systematic review. Computers & Education 123, 109–123 (2018)
12. Lewis, J.R., and Sauro, J.: The factor structure of the system usability scale. In: International conference on human centered design, pp. 94–103 (2009)
13. Lieberman, H., Paternò, F., Klann, M., and Wulf, V.: End-user development: An emerging paradigm. (2006)
14. Marfisi, I.: Teachers Paper-prototyping Augmented Reality Activities for Classroom Use. Mendeley Data (2022)
15. Mota, R.C., Roberto, R.A., and Teichrieb, V.: [POSTER] Authoring Tools in Augmented Reality: An Analysis and Classification of Content Design Tools. In: 2015 IEEE International Symposium on Mixed and Augmented Reality, pp. 164–167 (2015)
16. Nebeling, M., and Speicher, M.: The Trouble with Augmented Reality/Virtual Reality Authoring Tools. In: 2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct), pp. 333–337 (2018)
17. Radu, I.: Augmented reality in education: a meta-review and cross-media analysis. Personal and Ubiquitous Computing 18(6), 1533–1543 (2014)
18. Roopa, D., Prabha, R., and Senthil, G.: Revolutionizing education system with interactive augmented reality for quality education. Materials Today: Proceedings 46, 3860–3863 (2021)
19. Shams, L., and Seitz, A.R.: Benefits of multisensory learning. Trends in cognitive sciences 12(11), 411–417 (2008)
20. Wobbrock, J.O., Findlater, L., Gergle, D., and Higgins, J.J.: The aligned rank transform for nonparametric factorial analyses using only anova procedures. In: Proceedings of the SIGCHI conference on human factors in computing systems, pp. 143–146 (2011)
21. Xiao, M., Feng, Z., Yang, X., Xu, T., and Guo, Q.: Multimodal interaction design and application in augmented reality for chemical experiment. Virtual Reality & Intelligent Hardware 2(4), 291–304 (2020)
22. Yang, K., Zhou, X., and Radu, I.: XR-Ed Framework: Designing Instruction-driven andLearner-centered Extended Reality Systems for Education. (2020)
23. Ez-Zaouia, M.: Teacher-centered dashboards design process. In: 2nd International Workshop on eXplainable Learning Analytics, Companion Proceedings of the 10th International Conference on Learning Analytics & Knowledge LAK20 (2020)
24. Ez-Zaouia, M., and Lavoué, E.: EMODA: A tutor oriented multimodal and contextual emotional dashboard. In: Proceedings of the seventh international learning analytics & knowledge conference, pp. 429–438 (2017)
25. Ez-Zaouia, M., Marfisi-Schottman, I., Oueslati, M., Mercier, C., Karoui, A., and George, S.: A Design Space of Educational Authoring Tools for Augmented Reality. In: International Conference on Games and Learning Alliance, pp. 258–268 (2022)
26. Ez-Zaouia, M., Tabard, A., and Lavoué, E.: EMODASH: A dashboard supporting retrospective awareness of emotions in online learning. International Journal of Human-Computer Studies 139, 102411 (2020)

Utilisabilité d'un outil auteur immersif : une étude expérimentale pour la scénarisation de vidéos panoramiques interactives

Daniel Xuan Hien Mai[1], Guillaume Loup[1], Jean-Yves Didier[1]

[1] Université Paris-Saclay, Univ Evry, IBISC, 91020, Evry-Courcouronnes, France
danielxuanhien.mai@univ-evry.fr
guillaume.loup@univ-evry.fr
jeanyves.didier@univ-evry.fr

Résumé. La demande de formations à distance ainsi que le souhait de réduire les coûts d'apprentissage ont donné une forte impulsion au développement rapide d'environnements immersifs pour l'apprentissage humain. Une approche récente permet au formateur de concevoir son scénario pédagogique à l'aide d'un outil auteur pourvu d'une interface immersive et basé sur des vidéos panoramiques. Nous avons mené une étude pour évaluer l'utilisabilité de l'interface de Réalité Virtuelle (RV) et la comparer avec l'interface WIMP. Les résultats montrent (1) une meilleure corrélation entre les trajectoires des objets 3D positionnés par l'utilisateur et les entités ciblés dans la vidéo panoramique à l'aide d'une interface immersive ; (2) une différence significative dans le temps d'exécution des tâches entre les deux types d'interface ; (3) malgré le sentiment de cybermalaise, les formateurs sont toujours plus satisfaits et motivés lors de l'exécution de la tâche sur l'interface RV par rapport à l'interface WIMP.

Mots-clés : vidéos panoramiques, outil auteur, réalité virtuelle, interface WIMP, environnements immersifs, technique d'interaction.

Abstract. The need for distance education along with the desire to reduce learning costs has given a strong impetus to the rapid development of virtual immersive learning environments. Recently, a new approach has been developed to allow the trainer to construct his educational scenario with the help of an immersive authoring tool that is based on panoramic videos. We conducted this study to evaluate the usability of the Virtual Reality (VR) interface and to compare it with the WIMP interface. The results show (1) a better correlation between the trajectories of 3D objects positioned by the user and the entities targeted in the panoramic video using an immersive interface; (2) a significant difference in task execution time between the two interface types; (3) greater trainer satisfaction and motivation towards the VR interface compared to the WIMP interface despite the symptoms of cyber sickness.

Keywords: panoramic videos, authoring tool, virtual reality, WIMP interface, immersive environments, interaction technique.

1 Introduction

Le développement d'environnements immersifs présente de nombreux défis tant au niveau des processus que des outils de production [1]. Ces outils requièrent souvent des connaissances avancées du formateur pour atteindre les résultats souhaités. L'environnement immersif basé sur la vidéo panoramique est une solution qui simule des scénarios interactifs proches de la réalité. La vidéo panoramique peut être enrichie par l'ajout d'éléments interactifs afin d'améliorer l'expérience utilisateur [2]. Ils nécessitent une structure particulière, adaptée aux dimensions spatiales et temporelles [3]. Un outil auteur adapté assiste activement les formateurs dans la construction du contenu [4]. De nouvelles exigences pour les outils auteurs basés sur des vidéos panoramiques ont conduit au concept d'un outil exclusivement immersif. Il doit fournir aux formateurs un accès similaire à celui de l'apprenant afin d'obtenir rapidement un ensemble d'informations pour auto-évaluer les résultats de leur travail tout au long de la conception [5].

Cependant, il reste encore à déterminer si la nouvelle interface interactive en réalité virtuelle (RV) est différente en termes d'utilisabilité et de motivation par rapport à l'interface *Windows, Icons, Menus and Pointing* (WIMP) traditionnelle. Pour répondre à cette question, une expérimentation a été effectuée, selon un scénario défini, se déroulant sur les deux types d'interfaces.

2 État de l'art

La conception interactive qui complète la vidéo panoramique présente des défis non seulement techniques mais aussi de conception [6]. De plus, le sentiment d'immersion affecte l'efficacité de la navigation visuelle dans les interfaces WIMP et RV [7]. Une des principales tâches du concepteur sera de superposer un objet 3D à la vidéo panoramique. Pour assurer une cohérence entre ces deux entités, une trajectoire prédéterminée basée sur la chronologie de la vidéo panoramique [3] est définie. Plusieurs études ont été dédiées à l'outil auteur. T. Adão et al. ont réalisé une expérimentation pour évaluer l'utilisabilité de la solution de conception rapide d'objets [8]. Une autre expérimentation a évalué la continuité de l'intégration d'animations vidéo, d'objets 3D ainsi que de son 3D [9]. Pakkanen *et al*. ont proposé une comparaison de 3 modèles de techniques d'interaction en RV [10] utilisés pour contrôler la lecture de vidéo panoramique. Ils ont montré que les participants ont connu une réduction des nausées lors du deuxième passage. En tant qu'étude de cas d'outils auteurs de vidéos panoramiques, l'expérimentation de Coelho & Melo [5] est remarquable lorsqu'il s'agit d'évaluer l'utilisabilité de 3 types d'interfaces différents : WIMP, RV et tangible. La recherche a montré que des interfaces RV et tangibles ont un degré de satisfaction plus élevé que les interfaces WIMP.

Ces expérimentations ont montré que l'exécution de la même technique d'interaction sur différents types d'environnements ou d'interfaces aura des résultats d'utilisabilité différents. L'exigence de cohérence spatio-temporelle dans l'interaction est un élément caractéristique des environnements immersifs basés sur la vidéo panoramique. Il est

ainsi nécessaire de comparer une technique d'interaction spécifique aux outils auteurs sur vidéo panoramique, entre l'interface WIMP traditionnelle et l'interface RV, au travers d'une nouvelle étude expérimentale.

3 Hypothèses de travail

En ce qui concerne la différence dans des actions de changement du point de vue et d'enregistrement de trajectoire, notre première hypothèse était que cela conduirait à une différence d'utilisabilité entre deux types d'interface en termes de précision ainsi qu'en temps d'exécution. Notre deuxième hypothèse est que l'utilisation du casque RV n'augmente ni la charge mentale, ni les symptômes de cybermalaise de façon significative par rapport à une interface WIMP. Nous supposons que les participants utilisant l'interface RV sont plus motivés que ceux utilisant l'interface WIMP.

4 Méthode

4.1 Outil auteur immersif : Wixar [11]

Les deux versions PC et RV de Wixar ont la même interface utilisateur (UI) et modifient la technique d'interaction : la souris au lieu de la manette avec le laser, le clavier avec des boutons pour la rotation au lieu de tourner la tête avec le casque RV. Ces techniques d'interaction sont donc des interfaces homme-machines standards, aptes à la comparaison et n'affecteront pas négativement les résultats de l'expérimentation.

4.2 Participants, Scénarios et Procédure

Pour cette expérimentation, 30 participants ont été sélectionnés entre 20 et 55 ans, l'âge moyen étant de 30,7 ans. Les participants ont été répartis selon l'âge pour assurer un équilibre entre les deux groupes WIMP et RV. Les principales missions demandaient au participant de positionner et paramétrer un marqueur virtuel sur un poisson dans la vidéo panoramique. Chaque poisson possède une trajectoire différente avec une complexité croissante étape par étape : mouvement linéaire, léger ondulé, autour des pieds, autour de l'espace avec une accélération variable, autour verticalement.

4.3 Mesures et méthode d'analyse

Nous avons collecté les données relatives au comportement du participant : les mouvements, les opérations de création, suppression, déplacement d'objets et l'enregistrement de trajectoire du marqueur. Ces données ont été analysées pour évaluer l'efficacité (la précision de la manipulation) et l'efficience (le temps qu'il faut pour réaliser des tâches) de l'utilisabilité. Des questionnaires sont utilisés pour collecter des informations relatives à la démographie, au cybermalaise (SSQ) [12], à la satisfaction (SUS) [13] ainsi qu'à la motivation (SIMS) [14].

À un instant *t* dans la vidéo, le marqueur *m* a la position p_m et le poisson *p* a la position p_p. La distance $d_t(p_m, p_p)$ est la distance entre deux objets au temps *t*.

Notre objectif est de mesurer et comparer les variations et la stabilité des trajectoires enregistrées en tenant compte de la trajectoire ciblée. La distance relative entre deux périodes consécutives *t+1* et *t* a été calculée, c'est-à-dire $\Delta_{t+1,t} = |d_{t+1} - d_t|$.

5 Résultats

5.1 Effet de l'interface interactive sur l'efficacité

La différence moyenne (Δ) est calculée pour les tests d'efficacité sur les 5 missions dans des scènes de s1 à s5. L'efficacité est inversement proportionnelle au coefficient Δ, on peut donc voir (dans la figure 1) que le groupe des participants sur RV a enregistré des trajectoires plus stables que le groupe utilisant l'interface WIMP.

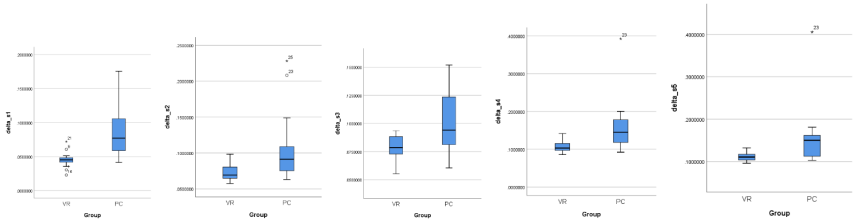


Fig. 1. Distribution des différences moyenne de trajectoire pour les deux groupes par séquence.

Le test de Student (sig.=0.016) sur Δs3 et les tests U de Mann-Whitney sur Δs1 (sig.=0.001), Δs2 (sig.=0.004), Δs4 (sig.=0.001) et Δs5 (sig.=0.005) montrent des différences significatives entre deux groupes.

5.2 Effet de l'interface interactive sur l'efficience

Concernant l'efficience, nous avons constaté que le groupe RV a terminé des tâches dans un temps relativement meilleur que le groupe WIMP (figure 2).

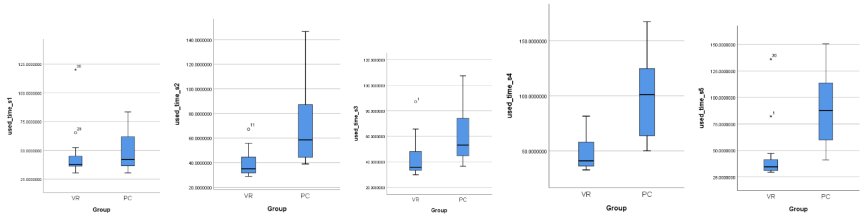


Fig. 2. Distribution des différences du temps d'exécution pour les deux groupes par séquence.

Des tests U de Mann-Whitney sur les données temporelles ont montré que seulement s1 (sig.=0.35) ne montrait aucune différence significative, les autres s2 (sig.=0.001), s3 (sig.=0.005), s4 (sig.=0.0001) et s5 (sig.=0.0002) montrent une différence significative.

5.3 Satisfaction, cybermalaise et motivation

Concernant les résultats du questionnaire F-SUS permettant d'évaluer le taux de satisfaction, il n'existe pas de différence significative entre le groupe RV (M=79.6, E.T.=11,7) et le groupe WIMP (M=76,8, ET=15,4). Même si les valeurs importantes des écarts types mettent en évidence une forte diversité des avis, l'interprétation des moyennes du SUS [15] permet de conclure que les deux systèmes mènent à un niveau de satisfaction acceptable par les utilisateurs.

Concernant les résultats des questions relatives à la nausée, le test de normalité par Shapiro-Wilk (sig.=0.009) mène à une évaluation U de Mann-Whitney (sig.=0.001) et a montré que les participants ayant porté le casque RV ont ressenti un cybermalaise significatif. De plus, le Test de Levene (sig.=0.447) et T-Test (sig.=0.023) sont confirmés par ceux des questions relatives aux troubles oculomoteurs plus important pour le groupe RV que pour le groupe WIMP.

Concernant la motivation intrinsèque, il existe une différence significative entre le groupe RV (M=23,8 et ET=2,9) et le groupe WIMP (M=19,7 et E.T.=5,4). Une seconde différence significative entre le groupe RV (M=8,9 et ET=4,6) et le groupe WIMP (M=13,9 et E.T.=7,1) concerne la régulation externe. Elles mettent en avant un sentiment d'autonomie plus important pour le groupe RV que pour le groupe WIMP.

6 Discussion

Concernant notre première hypothèse, l'interface RV s'est avérée plus efficace et plus efficiente. L'analyse a démontré que le suivi du poisson par un marqueur effectué sur l'interface RV est plus stable que celui effectué sur l'interface WIMP. La capacité à percevoir l'espace et surtout la vitesse de déplacement de l'objet à travers l'interface RV est meilleure que sur l'interface WIMP. La navigation spatiale, lorsqu'elle est combinée à un enregistrement de trajectoire simultané et ininterrompu, se traduit par de meilleures performances de l'interface RV en termes de temps et de précision.

D'après les résultats de l'analyse de cybermalaise, la nausée ainsi que le trouble oculomoteur étaient plus prononcés sur l'interface RV que sur l'interface WIMP. Le processus de l'expérimentation pour l'ensemble des 5 tâches est assez long, nous n'avons donc pas détaillé cet aspect de l'étude aussi profondément que l'expérimentation de Pakkanen et al. [10] où les participants répétaient la tâche pour évaluer leur capacité d'adaptation à l'environnement immersif à différents moments.

Les résultats des questionnaires de l'échelle de motivation situationnelle ont montré une différence significative dans la motivation intrinsèque et la régulation externe entre les deux groupes. Un plus grand sentiment d'autonomie dans le groupe RV se traduira par une meilleure motivation pour accomplir des tâches.

7 Conclusion

L'objectif de notre expérimentation était d'évaluer et de comparer l'utilisabilité de l'interface RV avec l'interface WIMP pour la conception de scénarios interactifs pour la vidéo panoramique. Les résultats sont meilleurs en termes de suivi de mouvement ainsi que de temps d'exécution sur l'interface RV par rapport à l'interface WIMP. Concernant la satisfaction, bien qu'il n'y ait pas de différence significative, elle était dans des limites acceptables. En termes d'utilisabilité, l'interface RV semble mieux adaptée grâce aux interactions relatives à la coordination spatio-temporelle. Bien que la RV ait des problèmes liés au cybermalaise, elle donne toujours une meilleure satisfaction et motivation aux formateurs en comparaison avec l'interface WIMP.

L'environnement immersif basé sur une vidéo panoramique interactive ne comprend pas seulement des objets interactifs spatio-temporels, mais également d'autres objets tels que du texte et du son. Par conséquent, la création de ces objets dans l'outil auteur

doit être évaluée plus en détail pour leur utilisabilité multiplateforme. À l'avenir, nous examinerons et évaluerons la pertinence du système pour différents types de scénarios, dans le but de proposer un modèle pour le développement d'un assistant de scénarisation afin d'aider le formateur dans le processus de construction.

References

1. Sundström, Y.: Game design and production: frequent problems in game development (2013).
2. Chambel, T., Chhaganlal, M. N., & Neng, L. A.: Towards immersive interactive video through 360 hypervideo. In Proceedings of the 8th international conference on advances in computer entertainment technology, pp. 1-2 (2011).
3. Mendes, P. R., Guedes, Á. L., Moraes, D. D. S., Azevedo, R. G., & Colcher, S.: An authoring model for interactive 360 videos. In 2020 IEEE International Conference on Multimedia & Expo Workshops (ICMEW) (pp. 1-6). IEEE (2020).
4. Khademi, M., Haghshenas, M., & Kabir, H.: A review on authoring tools. In Proceedings of the 5th International Conference on Distance Learning and Education, IPCSIT, vol. 12, pp. 40-44 (2011).
5. Coelho, H., et al.: Authoring tools for creating 360 multisensory videos—Evaluation of different interfaces. Expert Systems, 38(5), e12418 (2021).
6. Argyriou, L., et al.: Engaging immersive video consumers: Challenges regarding 360-degree gamified video applications. In 2016 15th international conference on ubiquitous computing and communications and 2016 international symposium on cyberspace and security (IUCC-CSS) (pp. 145-152). IEEE (2016).
7. Robertson, G., Czerwinski, M., & Van Dantzich, M.: Immersion in desktop virtual reality. In Proceedings of the 10th annual ACM symposium on User interface software and technology, pp. 11-19 (1997).
8. Adão, T. et al.: A rapid prototyping tool to produce 360 video-based immersive experiences enhanced with virtual/multimedia elements. Procedia computer science, 138, 441-453 (2018).
9. Choi, K., et al.: Interactive and Immersive Learning Using 360° Virtual Reality Contents on Mobile Platforms. Mobile Information Systems (2018).
10. Pakkanen, T., et al.: Interaction with WebVR 360 video player: Comparing three interaction paradigms. In 2017 IEEE Virtual Reality (VR), pp. 279-280. (2017).
11. Wixar, https://www.wixar.io/, last accessed 2023/05/11.
12. Kennedy, R. S., Lane, N. E., Berbaum, K. S., & Lilienthal, M. G.: Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. The international journal of aviation psychology, 3(3), 203-220 (1993).
13. Brooke, J.: SUS-A quick and dirty usability scale. Usability evaluation in industry, 189(194), 4-7 (1996).
14. Guay, F., Vallerand, R. J., & Blanchard, C.: On the assessment of situational intrinsic and extrinsic motivation: The Situational Motivation Scale (SIMS). Motivation and emotion, 24, 175-213 (2000).
15. Bangor, A., et al.: Determining what individual SUS scores mean: Adding an adjective rating scale. Journal of usability studies, 4(3), 114-123 (2009).

EIAH2023 : 11ème Conférence sur les Environnements Informatiques pour l'Apprentissage Humain
La conférence pluridisciplinaire francophone sur la conception et l'analyse des environnements numériques pour l'éducation et la formation
12-16 juin 2023 Brest (France)

Session de communications 2.2
Pédagogie et ludicisation

Le Chaudron Magique :

un jeu en Réalité Mixte pour l'apprentissage des fractions en autonomie

Sofiane Touel[1], Iza Marfisi-Schottman[1], Sébastien George[1], Lucas Hattab[2], Nicolas Pelay[2]

[1] LIUM, Le Mans Université 72085, le Mans, France
prenom.nom@univ-lemans.fr

[2] Plaisir Math Lab, Paris, France
prenom.nom@plaisir-maths.fr

Résumé. Les concepts mathématiques, et notamment la notion de fractions, sont complexes à acquérir. De plus, l'appréhension et l'anxiété que certains élèves peuvent ressentir envers les mathématiques ne facilitent pas leur apprentissage. Pour tenter de répondre à cette problématique, nous avons conçu un jeu pédagogique nommé Le Chaudron Magique*,* en collaboration avec des didacticiens de mathématiques, des enseignants et leurs élèves. Il s'agit d'une extension numérique du jeu existant et qui utilise la technologie de la Réalité Mixte pour faciliter l'apprentissage des fractions. Il propose plusieurs types d'exercice pour soutenir l'apprentissage des fractions sur la droite numérique, à travers un scénario ludique. Le jeu propose aussi des rétroactions, en temps réel, et des aides, pour rendre les élèves plus autonomes. Nous présentons l'expérimentation de ce jeu sur 228 élèves, en situation écologique de classe ainsi que les premiers résultats.

Mots-clés : Fraction, Réalité Mixte, jeux sérieux, anxiété de mathématique, apprentissage.

Abstract. Mathematical concepts, and especially the notion of fractions, are complex to acquire. Moreover, the apprehension and anxiety that some students may feel towards mathematics do not make their learning easier. To try to address this issue, we have designed an educational game called *The Magic Cauldron*, in collaboration with mathematics didacticians, teachers and their students. It is a digital extension of an existing game that uses Mixed Reality technology to facilitate the learning of fractions. It offers several types of exercises to support learning fractions on the number line, through a fun scenario. The game also provides real-time feedback and assistance to make students more autonomous. We present the experimentation of this game with 228 students, in an ecological classroom situation, as well as the preliminary results.

Keywords: Fraction, Mixed Reality, serious games, math anxiety, learning.

1 Introduction

L'enseignement des fractions représente une réelle difficulté [1]. Parmi les obstacles didactiques les plus fréquents se trouvent notamment le placement des fractions sur une droite graduée, la compréhension des fractions supérieures à un et la comparaison des fractions. Les évaluations internationales (PISA, TIMSS) pointent, année après année, les difficultés des élèves qui s'accroissent en France [2]. L'étude de Post et *al.* [3] montre que la complexité des fractions est notamment engendrée par la coexistence de plusieurs aspects de ce concept. Ils dénombrent cinq « personnalités » : une fraction peut être un opérateur, une partie d'un tout, un ratio, un quotient, une mesure et un décimal. Chacun de ces aspects est travaillé à travers différentes activités pédagogiques qui utilisent une grande multitude de représentations : sections d'une figure géométrique, paquets d'objets, droite numérique… Il semble y avoir un consensus sur le fait que la difficulté d'apprentissage des fractions repose sur la complexité du concept, la multiplicité de ces aspects et des représentations utilisées par les enseignants [1].

Un autre aspect, qui complique l'apprentissage des fractions, est l'anxiété que peuvent éprouver les élèves envers les mathématiques. D'après l'étude de Wigfield et Meece [4], l'anxiété précoce envers les mathématiques peut entraîner une augmentation du stress, des stratégies d'évitement et une répulsion des mathématiques, et plus tard, de toutes matières scientifiques, d'où l'importance de proposer des mesures d'atténuation de ce phénomène, dès le plus jeune âge.

Une des stratégies qui semble efficace pour réduire cette anxiété est d'introduire les mathématiques sous forme de jeu. L'entreprise *Plaisir Maths* a, par exemple, développé *l'Atelier des Potion*, un jeu didactique pour l'apprentissage des fractions qui montre des résultats très positifs en termes de motivation et d'apprentissage [5]. Un autre atout de ce jeu est le fait que les enfants manipulent des objets, ce qui facilite la comparaison des fractions de façon visuelle et intuitive [6]. Cependant, comme la grande majorité des ressources pédagogiques, ce jeu se focalise uniquement sur une forme de représentation (sections d'une figure géométrique) puisque les pièces de jeu sont des carrées, cercles, hexagones, et des rectangles découpés en demi, tiers, quarts…

Dans cet article, nous présentons le travail mené sur une extension numérique de ce jeu, nommé Le Chaudron Magique. Ce jeu combine, de façon simultanée, la représentation de sections d'une figure géométrique (pièces de jeu manipulées) avec la représentation de la droite graduée, grâce à la technologie de Réalité Mixte (RM). L'objectif est de faciliter la compréhension de cette nouvelle forme de représentation des fractions. La deuxième partie de cet article dresse un aperçu des recherches menées sur l'apprentissage des fractions et des atouts pédagogiques la RM. La section trois présente Le Chaudron Magique. La section quatre représente la méthodologie utilisée lors de l'expérimentation du jeu avec 228 élèves et les premiers résultats et analyses. Enfin, la dernière section propose des conclusions et perspectives.

2 État de l'art

La RM est une technologie qui se base sur la reconnaissance d'image. Elle permet non seulement aux utilisateurs de voir des informations virtuelles sur des objets réels (Réalité Augmentée), mais aussi d'interagir avec ces éléments virtuels en manipulant les objets réels. La RM présente de nombreux intérêts pédagogiques. Tous d'abord, cette technologie a l'avantage de s'appuyer sur la **manipulation d'objets** physiques, ce qui facilite l'apprentissage [7, 8]. Les augmentations (en Réalité Augmentée) facilitent également l'apprentissage de concepts compliqués [9].
En ce qui concerne les fractions, il est possible d'utiliser la RM pour reconnaitre les pièces de jeu et **afficher, de façon dynamique, une autre forme de représentation**. Les travaux du collectif du *Rational Number Project* montrent d'ailleurs que les enfants comprennent mieux les fractions quand elles sont enseignées avec plusieurs représentations et en faisant le lien entre celles-ci [10, 11]. Le principe est déjà utilisé dans plusieurs domaines comme la géométrie 2D et 3D [12, 13]. Ce principe a aussi donné de bons résultats pour enseigner les fractions [14].

En outre, la RM peut être utilisée pour afficher de l'aide contextualisé pour débloquer les élèves en difficulté [12]. Elle permet aussi de valider automatiquement les exercices [15] et ainsi **accroitre l'autonomie** des élèves afin que l'enseignant puisse se concentrer sur les élèves en difficulté. L'application *Le marathon des fractions* [15] propose notamment un mécanisme d'autocorrection et d'aide. Si les élèves font une erreur, ils perdent des points et sont invités à cliquer sur le bouton d'aide. Cette étude a démontré que les mécanismes de correction et d'aide ont été particulièrement efficaces pour leur donner plus d'autonomie aux élèves.

Enfin, la RM peut **accroître la motivation et l'engagement des élèves** [16] ce qui pourrait réduire l'anxiété liée aux mathématiques, comme démontré dans les projets *de magic boosed* [12] et le *marathon des fractions* [15].

3 Le Chaudron Magique

Le Chaudron Magique est un jeu en RM qui a été conçu en suivant une méthode de Design-Based Research avec des didacticiens des mathématiques, cinq enseignants et leurs classes, avec plusieurs cycles itératifs de conception et de développement [17]. Lors de ce travail avec les enseignants pilotes, il est apparu, qu'en France, le moment le plus délicat dans l'apprentissage des fractions est le passage des fractions, qui sont d'abord enseignées comme des sections d'une figure géométrique aux fractions sur la droite graduée (CM1, CM2, 6éme). Il semblait donc pertinent d'utiliser la RM et le jeu pour faciliter le passage entre ses deux formes de représentation.

Le Chaudron Magique propose un scénario de 30 à 45 minutes qui peut être faite en autonomie. Le scénario plonge l'élève dans la peau d'un apprenti sorcier qui découvre un chaudron magique. Malheureusement, un méchant diablotin sorcier dérègle le chaudron et l'enfant devra la rééquilibrer en complétant plusieurs activités. Comme le montre la Figure 1, les apprenants utilisent des pièces inspirées de l'*Atelier des Potions* qui sont reconnues par l'application. Il s'agit de formes géométriques qui représentent

quatre types d'animaux : raie, grenouille, araignée et serpent. Ils placent ses pièces sur une fiche, sous la tablette. Les élèves peuvent, à tout moment, demander de l'aide au petit fantôme, en cliquant sur le bouton « ? ». Ces aides ont été définies, par les enseignants, après avoir testé le jeu avec une classe. Un choix délibéré a été fait de ne pas utiliser le mot « droite graduée » qui est présenté dans le jeu comme le « magimètre ». Celui-ci est placé sous le chaudron et se met à jour quand les pièces de jeu sont posées dans le chaudron. En fonction des exercices, les valeurs affichées, en Réalité Augmentée, au début et à la fin du magimètre peuvent varier (ex : 0 à 1 ou 0 à 2) et il peut aussi être découpé en section (demi, tiers, quart…).

Le jeu est composé de trois niveaux (figure 2), qui proposent chacun un type d'activité pédagogique. Chaque niveau propose six exercices avec une difficulté croissante. Les élèves doivent réussir les exercices pour avancer dans le scénario.

Fig. 1. Matériel du jeu Le Chaudron Magique

Niveau 1 : Placer une fraction sur la droite avec l'aide des pièces de jeu. Chaque exercice possède des instructions du type : "Place le curseur du magimètre pour indiquer X de Y. Tu peux t'aider des pièces du jeu." où X indique la quantité demandée (sous forme de fraction) et Y le type d'ingrédient (ici un des quatre animaux). Quand l'élève place les pièces de jeu dans le chaudron, la valeur de la pièce apparait sur celle-ci, en Réalité Augmentée, et une ligne apparait sur le magimètre pour indiquer la bonne position. L'élève doit ensuite déplacer le curseur du magimètre au bout de cette ligne.

Niveau 2 : Placer une fraction sur la droite en la coupant en parts égales. Dans ce niveau, le curseur ne peut se déplacer que sur les graduations du découpage choisi par l'apprenant. Le principe de ce niveau est semblable au précédent : il faut placer le curseur sur la droite graduée pour la quantité demandée. Les consignes sont de la même forme que le niveau 1. Les différences étant que, cette fois-ci, placer les pièces dans le chaudron n'aidera pas l'apprenant, car cela n'entrainera pas de rétroaction. L'application montre en revanche les pièces correspondantes, en Réalité Augmentée, lorsque le

curseur est déplacé. Tous les découpages ne sont pas toujours proposés ce qui requière de travailler la notion d'équivalence des fractions.

Niveau 3 : Trouver les pièces de jeu qui correspondent à la fraction sur la droite. La dernière situation est la situation inverse au niveau 1. C'est-à-dire qu'à partir de la fraction indiquée sur le magimètre, l'apprenant doit placer la quantité d'ingrédients correspondante dans le chaudron. Le curseur sur la droite graduée est fixe. Là encore, le lien est fait entre la représentation sections d'une figure géométrique et la droite graduée.

Fig. 2. Interfaces du jeu Le Chaudron Magique

4 Expérimentation

L'expérimentation a pour objectif de valider ou invalider les hypothèses suivantes :

- H1 - L'utilisation du jeu Le Chaudron Magi*que* en RM peut soutenir le processus d'apprentissage des fractions.
- H2 - Les rétroactions du jeu Le Chaudron Magique en RM peuvent rendre les élèves plus autonomes (aide, correction et validation automatiques).
- H3 - L'utilisation du jeu Le Chaudron Magique en RM peut accroitre le plaisir et la motivation et contribuer à la baisse de l'anxiété des enfants concernant l'apprentissage des fractions.

4.1 Protocole expérimental

Afin de valider ces hypothèses, nous avons choisi de mettre en place une étude comparative entre un groupe test, qui utilise Le Chaudron Magique (GA), et un groupe contrôle, qui utilise des exercices classiques, sur papier, proposés par leurs enseignants (GB). Chaque groupe a suivi les étapes suivantes (figure 3): remplissage du pré-test de

connaissance, session d'apprentissage (soit classique soit avec Le Chaudron Magique), remplissage du post-test (identique au pré-test), session d'apprentissage avec la méthode d'enseignement qu'ils n'avaient pas encore testé (soit classique soit avec Le Chaudron Magique) et puis finir, remplissage d'un questionnaire final. Les deux groupes été équivalents (nombre de filles/garçons, niveau, motivation). De plus, les enseignants ont été impliqués dans la mise en place de l'étude et ont veillé à ce que les exercices classiques proposés au groupe contrôle soient représentatifs de la manière dont les fractions sont généralement enseignées.

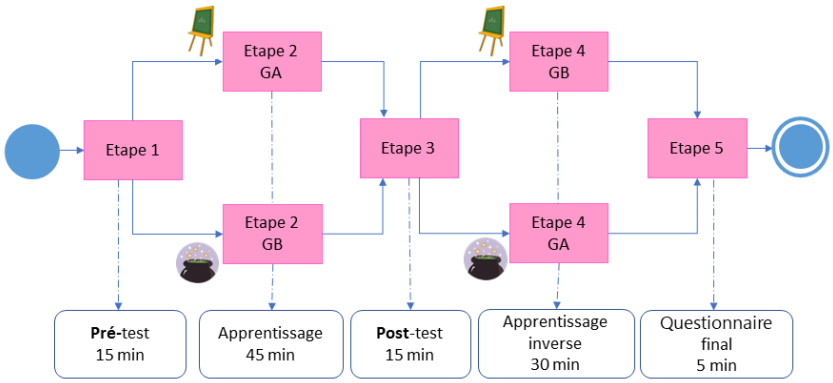


Fig. 3. Protocole expérimental

Les pré-test et post-test servent à mesurer la progression sur les compétences clés des fractions avant et après la session d'apprentissage surtout de la comparer entre groupes (H1). Ces tests sont composés de 12 questions : six pour évaluer les compétences sur les fractions inférieure ou égale à 1 et six pour évaluer les compétences sur les fractions supérieures à 1. Ils contiennent également des questions sur le gabarit *Child Math Anxiety Questionnaire* (CMAQ) [18] pour mesurer l'anxiété des enfants (H3). Les enfants doivent répondre à 12 questions à l'aide d'une échelle qui comporte un visage calme à droite et un visage stressé à gauche (figure 4). Ils doivent dessiner une flèche pour exprimer leur état émotionnel. Ces résultats sont reportés sur une échelle de 1 à 16.

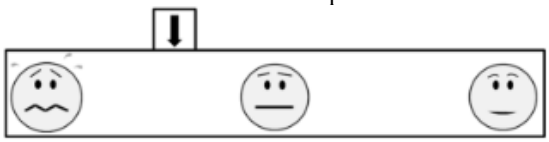


Fig. 4. Moyen de mesure de l'anxiété chez les enfants

Les sessions d'apprentissage ont été filmées, dans leur intégralité, pour compter le nombre de fois que les élèves ont sollicité l'aide de l'enseignant (H2). Pour finir, une fois que les deux groupes ont testé les deux méthodes d'enseignement, le questionnaire final permet de recueillir des informations sur la motivation des élèves et leurs comparaisons des deux sessions d'apprentissage (H3).

D'autres expérimentations complémentaires ont été menées, mais ne seront pas détaillées dans cet article par manque de place. Par exemple, pour avoir une vision sur l'impact du Chaudron Magique à plus long terme, le matériel de jeu a été laissé aux enseignants pendant une période complète (entre deux vacances scolaires).

4.2 Résultats

L'expérimentation a eu lieu entre janvier et juin 2022 sur huit classes : deux classes de CM1-CM2 à l'école Eugène Hairy, Laval (classe 1 : 23 élèves et classe 2 : 19 élèves), une classe de 5ème SEGPA du collègue Allain Gerbault, Laval (classe 3 : 19 élèves), cinq classes de 6ème au collège de la foresterie, Bonnetable (classe 4: 27 élèves, classe 5: 27 élèves, classe 6: 28 élèves, classe 7: 28 élèves, classe 8: 27 élèves) et une classe de 5e au même collège (classe 9: 30 élèves). L'expérience concerne 228 élèves au total.

Choix des indicateurs statistiques
Pour valider nos hypothèses, nous avons utilisé des tests statistiques adaptés aux données recueillies et aux types d'hypothèses testées. Le test de Wilcoxon est utilisé pour les hypothèses H1 et H3, concernant les différences de progression entre les groupes. Ce test non-paramétrique est approprié car il ne suppose pas de distribution normale des données, ce qui est souvent le cas pour les scores obtenus dans les tests éducatifs. De plus, il permet de comparer deux groupes indépendants (le groupe test et le groupe contrôle) sur une variable continue (la progression des élèves). Pour l'hypothèse H2, qui concerne la différence de répartition des réponses correctes/incorrectes entre les deux groupes, le test du Chi-carré (χ^2) est utilisé. Ce test permet de comparer la répartition de fréquences observées dans deux groupes indépendants pour une variable catégorielle.

Pré-traitement
Les données ont été filtrées pour enlever les absents au pré-test ou au post-test et les feuilles blanches (échantillon de 211 élèves). Nous avons aussi retiré une classe de 5éme SEGPA (classe d'enfant présentant de graves et durables difficultés scolaires) pour une question d'homogénéité, car le cours « classique » de l'enseignante était en fait composé de mini-ateliers de jeu de manipulation. Elle participait aussi activement à ces ateliers pour motiver ses élèves, ce qui aurait faussé les calculs d'autonomie. Les analyses suivantes ont donc été faites sur 195 élèves.

La fiabilité du questionnaire d'anxiété a été calculée avec le coefficient alpha de Cronbach pour les 12 questions. Le résultat de 0.738 suggère que les éléments ont une cohérence interne acceptable.

H1 - L'utilisation du jeu Le Chaudron Magique en RM peut soutenir le processus d'apprentissage des fractions.
Selon le test statistique de Wilcox, il y a une différence significative entre les résultats obtenus entre le pré-test et le post-test ($p = 2.281$ e-15). Autrement dit, tous les élèves ont acquis des compétences relatives aux fractions, qu'ils soient dans le GA ou le GB. La progression sur les compétences des fractions inférieures ou égales à 1 est aussi

bonne entre le GA et le GB (p = 0.6655). En revanche, la progression apparait moins bonne dans le GA par rapport au GB pour les compétences relatives aux fractions supérieures à 1 (p = 0.007392).

Nous nous attendions à avoir ce résultat, car la durée de l'expérimentation ne permet pas d'avoir une évolution remarquable. Cependant cette différence se manifeste uniquement dans les exercices qui traitent les fractions supérieures à 1, donc nous pensons que Le Chaudron Magique peut soutenir l'apprentissage des fractions chez les enfants. Nous estimons donc que notre hypothèse H1 est partiellement vérifiée.

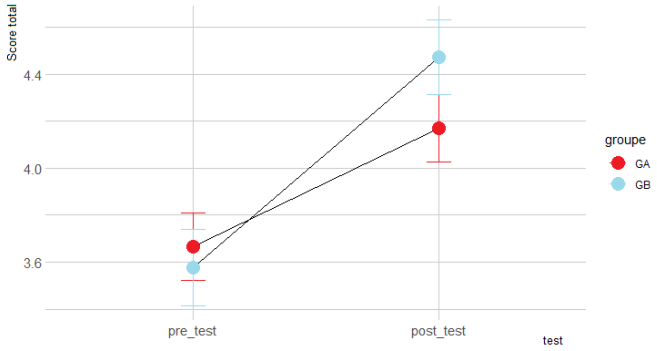


Fig. 5. Résultat aux tests de fractions du GA et GB (pre-test/post-test)

H2 - Les rétroactions du jeu Le Chaudron Magique en RM rendent les élèves plus autonomes.

Nous avons comparé les proportions de demande d'aide pour le GA et le GB, pour les trois niveaux de classe (cm1-cm2, 6éme et 5éme). Comme le montre la figure 5, les élèves demandent, en moyenne plus d'aide dans le GB. Selon le test statistique *khi-deux*, cette différence est significative (p = 3.528e-09). Grâce aux rétroactions en temps réel ainsi que le système d'aide à la demande, les élèves qui utilisent Le Chaudron Magique sont donc plus autonomes et sollicitent beaucoup moins leurs enseignants que les élèves qui participent à une séance classique, quel que soit leurs âges.

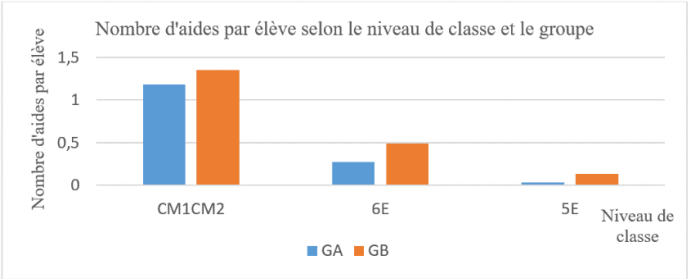


Fig. 6. Nombre d'aides moyens par élève du GA et GB

H3 - L'utilisation du jeu Le Chaudron Magique en RM peut accroitre le plaisir, la motivation et contribuer également à la baisse de l'anxiété des enfants concernant l'apprentissage des fractions.

Comme le montre la figure 6, l'anxiété a baissé entre le pré-test et post-test pour les deux groupes (échelle de 1 - stressé à 16 - calme). Cette différence est statistiquement significative. Cependant, il n'existe pas de différence statiquement significative entre GA et GB selon le test statistique de Wilcox (p = 0.01636). Le Chaudron Magique a donc contribué à la baisse de l'anxiété des enfants concernant l'apprentissage des fractions de façon significative, mais pas plus que l'activité classique, ce qui nous amène à analyser d'autres aspects afin de déterminer l'impact réel sur les différents sentiments ressentis lors de la séance telle que l'amusement, l'ennuie ou encore la motivation.

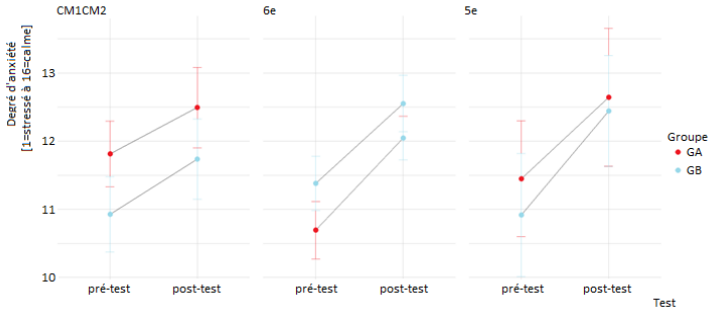


Fig. 7. Degré d'anxiété des élèves pendant le pre-test et le post-test par niveau

Nous remarquons aussi que les deux groupes sont moins stressé à la fin de l'expérimentation et aucune différence notable entre le GA et GB, d'après le questionnaire rempli par les élèves, cependant les mots que les élèves ont utilisés pour décrire les deux sessions d'apprentissage ainsi que la question sur leurs préférences, montre que ces derniers ont une préférence pour Le Chaudron Magique et sont plus motivé à utiliser ou réutiliser ce dernier.

La figure 8 montre le nuage de mot recueilli aux deux questions suivantes : « Écris 3 mots pour décrire la séquence d'exercices sur papier » et « Écris 3 mots pour décrire Le Chaudron Magique » posée dans le questionnaire final. Le top 10 des mots les plus répondus pour les activités classiques sont : 64 « bien », 42 « ennuyant », 39 « facile », 35 « long », 23 « dur », 22 « fraction », 19 « difficile », 18 « amusant », 16 « nul » et 10 « simple ». Le top 10 des mots les plus répondus pour Le Chaudron Magique sont : 83 « bien », 60 « amusant », 33 « facile », 31 « drôle », 28 « magique », 27 « super », 24 « rigolo », 24 « cool », 23 « chaudron » et 22 « fraction ». Nous remarquons que huit sur dix des mots les plus choisis par les élèves pour Le Chaudron Magique sont positifs, les deux autres sont neutres, contrairement à ceux des activités classiques, ce qui vient confirmer que les élèves ont plus de plaisir à jouer au Le Chaudron Magique qu'à participer à l'activité classique.

En ce qui concerne la motivation, les réponses sont aussi en faveur du Chaudron Magique. En effet, à la question « Que préfères-tu ? » les élèves ont choisis à 91% Le

Chaudron Magique contre 5% pour la séance papiers et 4% des élèves qui n'ont pas fourni de réponse ou ont donné des réponses erronées. Ceci vient confirmer le fait que les élèves ressentent plus de motivation à jouer au Chaudron Magique qu'à faire l'activité classique.

Nous pouvons donc conclure que l'utilisation du Chaudron Magique a permis d'accroitre le plaisir et la motivation et contribuer également à la baisse de l'anxiété des enfants concernant l'apprentissage des fractions ce qui valide l'hypothèse H3.

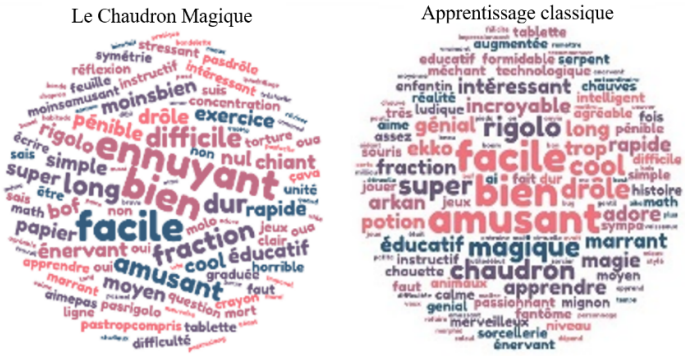


Fig. 8. Nuage de mots pour la session d'apprentissage classique et avec Le Chaudron Magique

4.3 Analyse et discussion

Bien que notre étude ait montré des résultats positifs en termes d'autonomie et de satisfaction des élèves, il est important de discuter des limites du jeu, notamment pour l'apprentissage des fractions supérieures à 1. Une piste d'amélioration pourrait consister à adapter le jeu pour mieux prendre en compte cet aspect, en intégrant par exemple des exercices spécifiques et des aides adaptées pour les fractions supérieures à 1.

De plus, l'autonomie des élèves a été mesurée en comptant le nombre de fois que ceux-ci levaient la main pour demander une question à l'enseignant. Cet indicateur présente toutefois des limites et des études futures pourraient explorer le temps passé à résoudre des problèmes de manière indépendante ou le nombre de problèmes résolus sans assistance.

Il serait également intéressant de développer une version du Chaudron Magique sans RM pour mesure l'impact du jeu et de la RM séparément.

5 Conclusion et perspectives

Les fractions demeurent une matière très mal aimée et difficile pour les élèves. De plus l'anxiété que ces derniers ressentent envers les matières des mathématiques, amplifie ce sentiment et décourage les élèves à essayer de comprendre et maitriser les fractions. Nous avons vu dans notre étude que les jeux pouvaient aider à atténuer ce sentiment envers les mathématiques. En nous basant sur *l'Atelier des potions,* qui a été conçu à

cet effet, nous avons donc développé une extension numérique qui s'appelle Le Chaudron Magique et qui intègre la technologie de Réalité Mixte (RM). Ce jeu a été conçu grâce à la collaboration entre des chercheurs informatiques, des didacticiens des mathématiques, des enseignants et leurs élèves. Il propose un parcours qui aide les élèves du cycle 2 (cm1, cm2 et 6éme) à mieux comprendre les fractions sur une droite graduée. Grâce à la RM, le jeu propose aussi un système d'aide contextualisé et d'autocorrection.

Nous avons testé notre jeu sur plus de 200 élèves. Nous avons remarqué que, durant cette séance de test, les élèves ont acquis des compétences sur les fractions, même si la différence dans une séance d'une heure ne montre pas d'amélioration significative par rapport aux cours classiques. Les élèves qui utilisent Le Chaudron Mag*ique* apprennent les fractions inférieure ou égale à 1 aussi bien que les élèves qui participent au cours classique. En revanche, ils apprennent légèrement moins bien les fractions supérieures à 1 avec Le Chaudron Magique. Nous remarquons aussi que les élèves qui ont joué au Chaudron Magique étaient plus autonome qu'en cours classique. Ceci est sans doute dû aux rétroactions, en temps réel, que nous permet d'afficher la MR ainsi que l'aide proposé dans le scénario du jeu. Enfin, Le Chaudron Magique a fait baisser le niveau de stress ressenti par les élèves, mais pas plus que le cours classique. En revanche, ils ont beaucoup plus apprécié jouer au Chaudron Magique.

Nous traitons actuellement les données récoltées sur les expériences à long terme avec Le Chaudron Magique et nous pensons que les résultats montreront des différences plus significatives entre les deux groupes, car il est difficile de démontrer l'apport du jeu en une seule séance. D'autres études complémentaires pourraient être menées pour affiner les résultats obtenus et explorer d'autres aspects susceptibles d'influencer l'impact du jeu sur l'apprentissage des élèves. Par exemple, il pourrait être intéressant d'étudier l'impact du jeu sur différents groupes d'élèves (selon leur niveau scolaire, profil).

References

1. Mercier, P.: Le passage de l'école primaire à l'école secondaire dans l'enseignement et l'apprentissage des fractions. Université Laval (2004).
2. Roditi, E., Martinez, S.: Programmes scolaires et apprentissage de la notion de fraction à l'école élémentaire. Quelques enseignements tirés de TIMSS. Éducation & formations. 94, 23–40 (2017).
3. Post, T., Behr, M., Harel, G., Lesh, R.: Rational numbers: Towards a semantic analysis–emphasis on the operator construct. Rational numbers: An integration of research. 13, (1993).
4. Wigfield, A., Meece, J.L.: Math anxiety in elementary and secondary school students. Journal of educational Psychology. 80, 210 (1988).
5. Pelay, N.: Jeu et apprentissages mathématiques: élaboration du concept de contrat didactique et ludique en contexte d'animation scientifique, (2011).
6. Pelay, N.: L'atelier des potions: Un jeu de société didactique et ludique pour apprendre les fractions. Revue de Mathematiques Pour l'école. 231, 6–11 (2019).
7. Fugate, J.M., Macrine, S.L., Cipriano, C.: The role of embodied cognition for transforming learning. International Journal of School & Educational Psychology. 7, 274–288 (2019).

8. Wilson, M.: Psychonomic bulletin & review. Six views of embodied cognition. 9, 625–636 (2002).
9. Da Costa, J., Szilas, N., Mueller, A.: Réalité augmentée pour l'apprentissage conceptuel en sciences: quels principes de conception pour les EIAH? Cas du dispositif DEAPE Learn en électromagnétisme. In: Actes de la 9ème Conférence sur les Environnements Informatiques pour l'Apprentissage Humain. pp. 181–192 (2019).
10. Cramer, K.A., Post, T.R.: Initial fraction learning by fourth-and fifth-grade students: A comparison of the effects of using commercial curricula with the effects of using the rational number project curriculum. Journal for Research in Mathematics Education. 33, 111–144 (2002).
11. Behr, M.J., Lesh, R., Post, T., Silver, E.A.: Rational number concepts. Acquisition of mathematics concepts and processes. 91, 126 (1983).
12. Andrea, R., Lailiyah, S., Agus, F., Ramadiani, R.: "Magic Boosed" an elementary school geometry textbook with marker-based augmented reality. TELKOMNIKA (Telecommunication Computing Electronics and Control). 17, 1242–1249 (2019).
13. Liou, H.-H., Yang, S.J., Chen, S.Y., Tarng, W.: The influences of the 2D image-based augmented reality and virtual reality on student learning. Journal of Educational Technology & Society. 20, 110–121 (2017).
14. Radu, I., McCarthy, B., Kao, Y.: Discovering educational augmented reality math applications by prototyping with elementary-school teachers. In: 2016 IEEE Virtual Reality (VR). pp. 271–272. IEEE (2016).
15. Palaigeorgiou, G., Tsolopani, X., Liakou, S., Lemonidis, C.: Movable, resizable and dynamic number lines for fraction learning in a mixed reality environment. In: International Conference on Interactive Collaborative Learning. pp. 118–129. Springer (2018).
16. Chen, Y.: Effect of mobile augmented reality on learning performance, motivation, and math anxiety in a math course. Journal of Educational Computing Research. 57, 1695–1722 (2019).
17. Marfisi-Schottman, I., Touel, S., George, S.: Designing a Mixed Reality Extension for an Educational Board Game on Fractions. International journal of virtual reality. 21, (2021).
18. Ramirez, G., Gunderson, E.A., Levine, S.C., Beilock, S.L.: Math anxiety, working memory, and math achievement in early elementary school. Journal of Cognition and Development. 14, 187–202 (2013).

Détermination de profils relatifs à la mobilisation de schème lors de la résolution de puzzles de programmation

Marielle Léonard[1,2], Maxime Bouton[1] et Yvan Peter[1]

[1] Univ. Lille, CNRS, Centrale Lille, UMR 9189 CRIStAL, F-59000 Lille, France
[2] Univ. Lille, ULR 4354 - CIREL - Centre Interuniversitaire de Recherche en Éducation de Lille, F-59000 Lille, France
prénom.nom@univ-lille.fr

Résumé. En nous appuyant sur la théorie des champs conceptuels développée par Vergnaud, et plus précisément sur le concept de schème, nous étudions les compétences d'élèves de collège à résoudre des puzzles de programmation en langage Scratch. À cette fin, nous construisons une combinaison d'indicateurs qui nous permet de déterminer automatiquement, à partir de traces d'interaction, des profils relatifs à l'existence de la mobilisation d'un schème opérationnel pour résoudre un puzzle de programmation.

Mots clés : Programmation par blocs , Apprentissage de la programmation , Puzzle de programmation , Schème , Analyse de traces

Résumé. This article presents our study of secondary school pupils fluency during programming puzzles solving using the Scratch language. We analyse their interaction traces through the scope of Vergnaud's theory of conceptual fields (scheme concept). We built a combinaison of indicators that can determine fluency profiles automatically.

Mots-clés : Block-based programming , programming learning , Puzzle , Scheme , Learning analytics

1 Introduction

Depuis 2016 en France, la programmation informatique est présente dans les programmes scolaires de l'ensemble de la scolarité primaire et secondaire. Au collège, c'est la programmation par blocs avec des langages comme Scratch qui est préconisée. Pour ce niveau collège, nous nous intéressons à la résolution de problèmes de programmation de type *puzzle* dans des micromondes [7]. Il s'agit de problèmes fermés, où la tâche à réaliser est prescrite et pour lesquels une solution proposée peut être évaluée de manière exacte, comme valide ou non.

Notre objectif est d'étudier les compétences des élèves à résoudre ce type de problème à partir de leurs traces d'interaction avec l'environnement de programmation. Nos questions de recherche sont formulées de la manière suivante :

— RQ1 : Quels indicateurs construits à partir de traces d'interaction rendent compte de l'existence de compétences à résoudre des puzzles dans un environnement de programmation par blocs ?
— RQ2 : Quels profils relatifs au degré de compétence à résoudre des puzzles dans un environnement de programmation par blocs peut-on définir à partir de ces indicateurs ?

De nombreuses acceptions du terme de compétences existent [3]. Pour notre part, nous nous appuyons sur des éléments de la théorie des champs conceptuels de Vergnaud, que nous présentons dans une première partie. [11]. Nous complétons par la consultation de travaux relatifs à l'analyse de traces d'interaction. Nous décrivons ensuite notre cadre expérimental ainsi que la construction de nos outils d'analyse : visualisation, indicateurs. Nous présentons les résultats obtenus avant de conclure et d'aborder les perspectives.

2 Compétence et schème dans la théorie des champs conceptuels

Les puzzles de programmation relèvent d'une approche constructiviste de l'apprentissage, développée par Piaget [8]. Cette approche a été reprise et approfondie notamment par Gérard Vergnaud, auteur de la théorie des champs conceptuels [11]. Cette théorie vise à comprendre la conceptualisation, en particulier dans le cas des activités cognitives complexes, dont la programmation informatique fait partie.

Selon Vergnaud, « Sans conceptualisation, il n'y a pas d'activité opératoire possible et pas de compétence »[10]. Par compétence, Vergnaud désigne « "la forme opératoire de la connaissance", qui permet de faire et de réussir »[13]. Pour Vergnaud, le concept de compétence présente une limite, dans la mesure où il « renvoie d'abord au résultat de l'activité et insuffisamment à l'organisation de l'activité elle-même. »[12]. Malgré cela, il convoque ce concept à plusieurs reprises, mentionnant par exemple que « A est plus compétent que B, s'il s'y prend d'une meilleure manière. Le comparatif "meilleure" suppose des critères supplémentaires : rapidité, fiabilité, économie, élégance, compatibilité avec la manière de procéder des autres, etc... »[13] Concernant l'organisation de l'activité, elle est prise en charge par le concept de schème, défini comme une « organisation invariante de la conduite pour une classe donnée de situations » [11].

Vergnaud catégorise les situations (ou tâches) en deux classes :

— des classes de situations pour lesquels le sujets dispose des compétences nécessaires au « traitement relativement relativement immédiat de la situation ». Dans ce cas, la conduite du sujet est largement automatisée et organisée autour d'un schème unique.

— des classes de situations pour lesquels le sujet ne dispose pas de toutes les compétences nécessaires. Dans ce cas, le sujet entre dans un processus d'accommodation, lors duquel des schèmes sont essayés, déconstruits, réorganisés.[11, p. 136]

Nous retenons que la durée nécessaire au traitement de la situation est la variable déterminante pour distinguer ces deux classes de situation et que les compétences sont associées à la mobilisation d'un schème opérationnel pour traiter la situation.

Dans notre contexte de résolution de puzzles de programmation, nous cherchons donc un ou plusieurs indicateurs construits à partir de traces d'interaction qui permettent de rendre compte de la durée irréductible à l'exécution de la tâche (traitement immédiat révélateur de la mobilisation d'un schème unique) et du temps éventuellement nécessaire à un processus d'accommodation. Notre objectif est alors de déterminer automatiquement des profils, que nous appelons profils relatifs à la mobilisation de schème, fondés sur le traitement immédiat ou non du puzzle de programmation.

3 Analyse des traces d'interaction

Nous relevons des travaux sur lesquels nous appuyer pour construire des indicateurs à partir de traces d'interaction collectées lors d'une activité d'apprentissage et déterminer des profils par rapport au traitement d'une situation.

Pelanek et al. dressent un panorama des aspects à prendre en considération lors de la modélisation d'un domaine et de la modélisation d'un processus d'apprentissage, en fonction du contexte et de l'objectif de la modélisation [5]. Dans le tableau de synthèse présenté par ces auteurs, nous relevons qu'en plus de la validité des réponses, le temps nécessaire pour résoudre la tâche et l'historique des réponses soumises sont indiqués pour mesurer l'aisance (*fluency*) lors d'un processus d'apprentissage.

Dans un autre article, Pelanek explore le temps pour résoudre une tâche, associé à la nature des réponses fausses [6]. Il construit un indicateur pour normaliser le temps de réponse par rapport à la difficulté de la tâche. Cet indicateur reste cependant dépendant de l'échantillon considéré. Par ailleurs, il catégorise les erreurs de manière binaire en *commune* et *non commune*. À cette fin, il définit un seuil d'occurrences par rapport au total des réponses fausses. En combinant ces deux indicateurs, il établit quatre *profils de performance* : réponse correcte, réponse fausse dans un temps très court, erreur commune sans que le temps soit court, autre erreur.

Jiang et al. établissent également, en utilisant une méthode de *clustering*, des profils de comportement lors de la résolution d'un problème de programmation [4]. Ils caractérisent quatre profils à partir des programmes intermédiaires soumis : ceux qui abandonnent rapidement (*quitters*), ceux qui sont loin d'une solution valide lors des premiers essais, s'en approchent plus ou moins sans aboutir (*approachers*), ceux qui sont loin d'une solution valide lors

des premiers essais puis s'en approchent jusqu'à trouver (*solvers*), ceux qui soumettent un programme valide dès les premiers essais (*knowers*)

De ces travaux nous retenons l'idée de combiner deux indicateurs pour caractériser le traitement immédiat ou non d'une situation, et donc l'existence d'un schème opérationnel. Ces deux indicateurs sont le temps passé sur la tâche et nombre de tentatives (prise en compte des états intermédiaires). Nous retenons aussi l'idée de normaliser l'indicateur construit par rapport à des variables que nous souhaitons écarter temporairement de notre analyse.

4 Cadre expérimental

Les traces d'interaction que nous analysons sont issues de 658 sessions d'élèves de collège sur les parcours du concours en ligne de programmation Algorea[3] de l'année 2022. Chaque parcours est composé de 6 problèmes déclinés en 4 versions de difficulté croissante, ce qui fait 24 puzzles par parcours. Nous disposons donc de traces d'interaction de sujets avec l'environnement de programmation pour 144 puzzles différents (3 tours du concours dans 2 catégories), qui recouvrent les notions de base de l'algorithmique (séquence d'instructions, boucle, structure conditionnelle, variable, procédure sans paramètre). La figure 1 montre l'interface de programmation avec un exemple de puzzle.

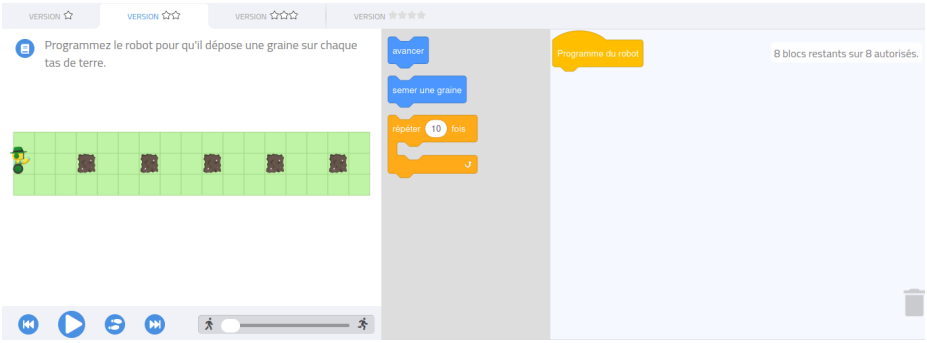


Fig. 1. Puzzle dans l'environnement de programmation Algorea

Lorsque l'utilisateur exécute un programme, celui-ci est évalué automatiquement par le système et un enregistrement est généré. L'évaluation du programme prend en compte les états intermédiaires et l'état final de la grille (le robot virtuel a-t-il bien rempli la mission qui lui est assignée en respectant les contraintes du problème ?) et les contraintes au niveau de

3. Concours de programmation en ligne organisé annuellement par l'association France-ioi (https://algorea.org)

l'édition du programme (limite du nombre d'instructions). Chaque enregistrement collecté contient l'identifiant de session, l'identifiant du problème, la version du problème, le programme exécuté, le résultat de l'évaluation du système (correct ou non) et l'horodatage associé à la soumission. Des enregistrements sont générés en plus pour certains événements importants comme un changement de puzzle. Après nettoyage des données, nous disposons de 74727 enregistrements utiles pour notre étude qui correspondent à 12340 confrontations d'un sujet avec un puzzle.

En complément, nous disposons également d'enregistrements vidéo de sessions pour une vingtaine de sujets (enregistrement de l'écran et de la voix). Ceux-ci nous permettent de disposer d'informations complémentaires pour appuyer nos choix méthodologiques.

5 Construction des outils pour l'analyse des données

5.1 Génération d'une visualisation sous forme de frise chronologique

Afin de disposer d'un premier outil pour explorer nos données, nous générons une vue synthétique de la session du sujet sous forme de frise chronologique (Figure 2). Ce type de visualisation est présent dans la littérature, mais plutôt pour l'étude du processus de résolution d'un problème particulier ([9], [2]).

Fig. 2. Visualisation de la session d'un sujet sous forme de frise chronologique (zoom sur les puzzles 4 et 5 du parcours, entre les $6^{\text{ème}}$ et $18^{\text{ème}}$ minutes)

L'axe horizontal est un axe temporel qui porte les différents évènements survenant pendant la session : changements de puzzle et exécutions. Chaque problème est représenté par une couleur dédiée, chaque version de problème (puzzle) par une nuance et une hauteur : plus clair et plus bas pour la version la plus facile jusqu'à plus foncé et plus haut pour la version la plus difficile de chaque problème. Chaque exécution d'un programme invalide est représentée par un trait noir, chaque soumission d'un programme valide par un trait blanc.

5.2 Construction d'indicateurs pour la détermination de profils

En nous appuyant sur le cadre d'analyse de Vergnaud présenté dans la section 2, nous cherchons à déterminer automatiquement, à partir des traces d'interaction, si le sujet réussit un puzzle de programmation de manière

experte, en mobilisant un schème opérationnel (validation rapide et quasiment sans se tromper), ou s'il s'engage dans un processus d'accommodation.

Pour cela, nous avons besoin de calculer le temps passé sur chaque puzzle et le nombre d'essais réalisés avant d'aboutir à une solution valide ou à un abandon. En utilisant ces données, nous déterminons un seuil en temps et en nombre d'essais au-delà duquel nous considérons qu'un processus d'accommodation est engagé, qu'il aboutisse à une résolution du problème (profil accommodation) ou non (profil échec).

Nous fixons le seuil en nombre d'essais à trois, ce qui inclut la nuance d'un schème opérationnel qui fait l'objet d'une ou deux erreurs de mise en œuvre, et qui est ajusté aussitôt (profil ajustement). En effet, nous avons remarqué que dans ce cas, le programme soumis est corrigé très rapidement, puis à nouveau exécuté et validé par le système.

Comme argumenté dans la section 2, la vitesse de résolution du problème est essentielle pour déterminer le degré de compétence d'un sujet face à une situation. Cette vitesse dépend du nombre de blocs à déplacer pour obtenir le programme solution et de la dextérité du sujet. Un sujet plus jeune peut disposer d'un schème opérationnel mais une moindre dextérité avec la souris induit un temps de résolution plus long. Notre objectif est de construire un indicateur qui ne dépende plus ni du nombre de blocs à déplacer pour obtenir la solution de référence, ni de la dextérité du sujet.

Dans une première étape, nous calculons un temps moyen individuel pour placer un bloc dans l'éditeur (t_{bloc}). Le calcul de t_{bloc} prend seulement en compte la première soumission pour chaque puzzle abordé. En effet, cette première soumission correspond au schème mobilisé en première intention pour traiter la situation, qu'il soit opérationnel ou non. À l'inverse des soumissions suivantes où les comportements sont plus divers, nous observons (sessions vidéos en appui) très peu de latence entre les déplacements de blocs lors de cette première soumission.

$$t_{bloc} = \frac{\sum_i t_{prem}(i)}{\sum_i N_{prem}(i)} \tag{1}$$

t_{bloc} est calculé selon la formule 1 où $\sum_i t_{prem}(i)$ est la somme sur tous les puzzles du parcours des durées des premiers essais et où $\sum_i N_{prem}(i)$ est le nombre de blocs déplacés lors de l'ensemble des premiers essais.

Nous construisons ensuite un indicateur de rapidité r_i attaché à un puzzle i (formule 2). Celui-ci prend en compte le nombre de blocs à déplacer, en nous basant sur le nombre de blocs de la solution de référence $N_{ref}(i)$ pour ce puzzle i et sur t_i qui est le temps passé sur ce puzzle. Dans le cas d'un expert qui soumet la solution de référence en un seul essai, cet indicateur correspond au temps moyen nécessaire pour placer un bloc. À ce stade, cet indicateur ne tient pas encore compte de la dextérité du sujet.

$$r_i = \frac{t_i}{N_{ref}(i)} \tag{2}$$

Nous calculons enfin un indicateur d'accommodation a_i attaché à la résolution du puzzle i (formule 3). Pour cela, nous centrons l'indicateur r_i autour de 0 en prenant l'écart entre r_i et t_{bloc} et nous le réduisons en divisant par le temps moyen mis pour placer un bloc. Nous obtenons ainsi un indicateur qui est indépendant de la dextérité du sujet et du nombre de blocs de la solution de référence.

$$a_i = \frac{r_i - t_{bloc}}{t_{bloc}} \tag{3}$$

Si cet indicateur a_i est négatif, cela signifie que le sujet a résolu plus rapidement le puzzle i que la moyenne de ses premières soumissions. Dans ce cas, il y a absence d'accommodation, le sujet mettant en oeuvre un schème unique si l'on se réfère au cadre d'analyse de Vergnaud. Plus cet indicateur est élevé, plus le sujet a passé de temps d'accommodation sur le puzzle i relativement à sa dextérité et au nombre de blocs nécessaires pour résoudre ce puzzle. Combiné au nombre d'essais et à la validité de la dernière réponse soumise, cet indicateur d'accommodation nous permet, après fixation d'un seuil, de déterminer automatiquement des profils relatifs à la mobilisation de schèmes lors de la confrontation de sujets avec des puzzles de programmation.

6 Résultats et discussion

Fig. 3. Identification de profils sur la frise chronologique d'une session (zoom)

La visualisation de la session d'un sujet sous forme de frise chronologique nous permet d'observer plusieurs profils construits à partir de la distinction de Vergnaud à propos du traitement immédiat ou non d'une situation (Figure 3). Un profil est associé à un couple sujet/puzzle. Différents profils se succèdent pour un même sujet au cours d'une session. :

— Profil **expert** : le sujet résout le puzzle très rapidement en un seul essai. La résolution du puzzle est organisée par un schème unique, opérationnel dans la situation.
— Profil **ajustement** : le sujet résout le puzzle rapidement après une ou deux soumissions ratées. Il s'organise autour d'un schème opérationnel pour traiter la situation, mais avec des approximations lors de la mise en œuvre.

— Profil **accommodation** : le sujet passe un temps conséquent sur le problème et fait plusieurs soumissions ratées avant de soumettre un programme valide. Il ne dispose pas de schème directement opérationnel pour traiter la situation. Il entre dans un processus d'accommodation qui aboutit positivement.
— Profil **échec** : le sujet est en échec sur le problème. Comme pour le profil précédent, le sujet entre dans un processus d'accommodation, mais celui-ci ne permet pas d'aboutir à un programme valide pour résoudre le puzzle.

Nous identifions par ailleurs deux autres profils, non visibles sur cet exemple de frise :

— Profil **passé** : le sujet ouvre brièvement un puzzle sans lancer d'exécution de programme.
— Profil **indéterminé** : le sujet résout le puzzle en un seul essai mais avec un temps long, qui ne nous permet pas de conclure sur la mobilisation d'un schème opérationnel. Nous observons plusieurs cas sur nos enregistrements vidéos où le sujet suspend momentanément sa recherche pour faire autre chose, mais aussi des cas où le sujet manipule longuement des blocs dans l'éditeur avant de lancer une exécution et des cas où le sujet consulte le tutoriel disponible.

La question est, à ce stade, de définir un seuil pour l'indicateur d'accommodation construit dans la section 5.2. En effet, ce seuil détermine la répartition entre les profils expert et indéterminé d'une part, ajustement et accommodation d'autre part. Nous visualisons dans ce but la courbe de la fréquence des profils expert et ajustement en fonction du seuil considéré pour l'indicateur d'accommodation (Figure 4).

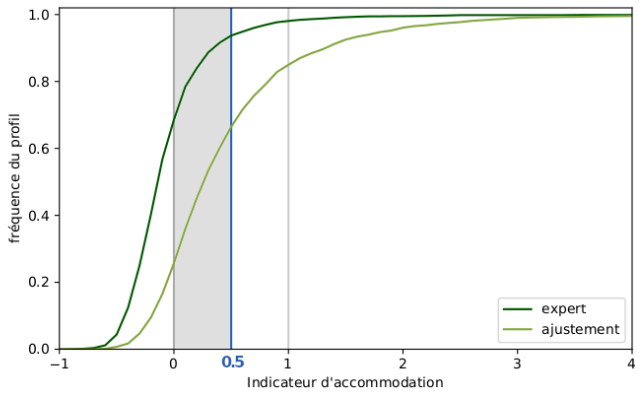


Fig. 4. Fréquence des profils expert (réussite en une seule tentative) et ajustement (réussite en deux ou trois tentatives) en fonction du seuil retenu pour l'indicateur d'accommodation

De cette courbe, nous déduisons un intervalle de valeurs dans lequel nous situons le seuil recherché et concentrons nos analyses. La valeur 0 pour l'indicateur d'accommodation équivaut à une résolution avec un temps égal au temps moyen des premiers essais du sujet. Lorsque la valeur de l'indicateur d'accommodation vaut 0,5 cela signifie que le sujet met 1,5 fois plus de temps à résoudre le puzzle que le temps moyen de ses premiers essais.

Nous avons fait varier le seuil entre 0 et 0,5 et étudié l'incidence sur la répartition des profils. Nous avons finalement retenu un seuil à 0,5. La fixation du seuil à cette valeur comporte certes une part d'arbitraire. Elle permet cependant un équilibre satisfaisant entre :

— le résidu des validations en 1 essai (profil indéterminé) qui ne sont pas catégorisées dans le profil expert
— une distinction entre les profils ajustement et accommodation concordante dans une majorité de cas avec nos observations sur les enregistrements vidéo

En effet, nous avons confronté pour chaque couple sujet/puzzle le profil déterminé automatiquement à partir d'un seuil à 0,5 avec le profil attribué suite à une analyse de l'enregistrement vidéo (pour 73 cas sur 1277, soit 5,7%, le profil reste indéterminé ou non pris en charge par la catégorisation). Nous obtenons une concordance de 98% pour le profil échec, 92% pour les profils expert et accommodation, 70% pour le profil ajustement. Le profil indéterminé calculé à partir du seuil correspond, sur la base de l'analyse vidéo, pour 29% à un profil expert, 6% à un profil ajustement, 50% à un profil accommodation et pour 15% à un cas non pris en charge par la présente catégorisation (notamment suspension temporaire de l'activité).

En considérant ce seuil à 0,5, nous visualisons un histogramme des profils identifiés selon la valeur de l'indicateur d'accommodation (Figure 5). Ainsi, sur les 12340 confrontations de sujet à un puzzle de programmation, nous obtenons :

— 2819 occurrences (soit 22,8%) de profil *expert* : puzzle résolu en un seul essai et $a_i \leq 0.5$
— 1728 occurrences (soit 14,0%) de profil *ajustement* : puzzle résolu en 2 ou 3 essais et $a_i \leq 0.5$
— 4049 occurrences (soit 32,8%) de profil *accommodation* : puzzle résolu en 2 ou 3 essais et $a_i > 0.5$ ou résolution en plus de 3 essais.
— 2407 occurrences (soit 19,5%) de profil *échec* : puzzle non résolu, toutes les soumissions sont invalides.
— 1067 occurrences (soit 8,6%) de profil *passé* : puzzle ouvert mais pas de programme soumis.
— 270 occurrences (soit 2,2%) de profil *indéterminé* : un seul essai mais $a_i > 0.5$.

Cette étude montre qu'une part significative des réussites aux puzzles de programmation ne relèvent pas d'une résolution experte. En effet, parmi les 71,8% de confrontations d'un sujet à un puzzle qui aboutissent à une réussite (programme valide), seulement 51,3%, soit un peu plus de la moitié, relève d'un traitement relativement immédiat de la situation (profils expert et ajustement).

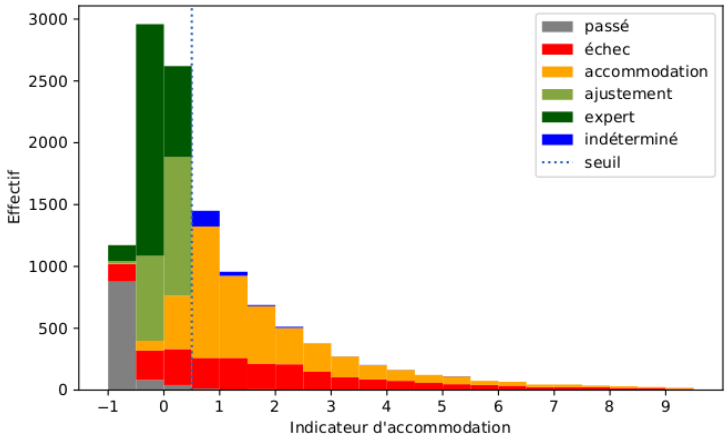


Fig. 5. Histogramme des profils avec un seuil de l'indicateur d'accommodation à 0.5 pour la détermination des profils expert et ajustement

Dans ce cas, la résolution révèle l'existence des compétences en jeu, et l'activité du sujet est organisée selon un schème unique. Ce résultat nous amène donc à inciter à la vigilance sur le fait d'utiliser un résultat brut (seulement réussite ou échec) lié à une série de puzzles pour évaluer des compétences en programmation, surtout lorsque le nombre d'exécutions du programme n'est pas limité. En revanche, la combinaison de la réussite au puzzle, du nombre d'essais et de l'indicateur d'accommodation proposé dans cet article, permet de rendre compte de manière plus fiable de la maîtrise des compétences en programmation sous-jacentes, à la condition que les compétences nécessaires à la résolution de chaque puzzle aient été correctement identifiées.

7 Conclusion & Perspectives

Dans ce travail nous avons cherché à déterminer le profil d'un sujet relatif à la mobilisation de schème lors de sa confrontation à un puzzle de programmation en langage Scratch. Notre contribution porte sur la mise en relation d'éléments de la théorie des champs conceptuels de Vergnaud avec l'analyse de traces d'interaction dans le but de quantifier le caractère immédiat ou non du traitement d'une situation et de caractériser automatiquement des profils dans le contexte de la résolution de puzzles de programmation en langage Scratch. À cette fin, nous avons calculé un indicateur d'accommodation attaché à un couple sujet/puzzle indépendant de la dextérité du sujet et de la solution de référence, que nous avons combiné avec le nombre d'essais et la validité ou non des programmes exécutés.

Associé à une identification précise des compétences en jeu dans chaque puzzle, cette détermination de profil sur la base de l'indicateur

d'accommodation et du nombre d'essais peut aider à automatiser la validation de compétences en programmation par blocs. Par ailleurs, nous avons implémenté dans un travail précédent un modèle d'apprentissage machine permettant une prédiction binaire (réussite / échec) de la réussite d'un sujet à un puzzle [1]. Nous avons adapté ce modèle pour la prédiction de nos profils avec une précision similaire au résultat précédent. Cette prédiction ouvre des perspectives de rétroactions personnalisées et d'adaptations de parcours.

Dans la mesure où nous n'avons pris en compte que la validité des réponses, le temps de réponse et le nombre d'essais, indépendamment du contenu des puzzles, ces résultats pourraient probablement être généralisés à des tâches pour lesquelles il est possible de définir une action élémentaire (ici placer un bloc dans l'éditeur). Nous notons que Pelanek a déjà entamé ce type de généralisation à des données de différentes natures [6].

Néanmoins, une limite de la démarche proposée dans cet article est qu'elle ne permet pas d'investiguer l'organisation invariante de l'activité, centrale dans le concept de schème. Nous déduisons seulement l'existence d'un schème unique organisateur de la conduite dans le cas des résolutions expertes, en nous fondant sur le cadre théorique de Vergnaud. Or selon l'auteur, « Le concept de compétence appelle une analyse de l'activité en termes de "schèmes". »[14]. Le présent travail n'est donc qu'une étape dans l'analyse de l'activité du sujet confronté à des puzzles de programmation. Se posent notamment des questions sur l'existence d'un apprentissage suivant la manière dont le sujet opère : est-ce que le sujet analyse le résultat de chaque exécution de son programme et en déduit les modifications à apporter pour aller vers une solution valide, procède-t-il par tâtonnement en exécutant une succession de programmes légèrement modifiés en espérant tomber par chance sur une solution valide, existe-t-il d'autres stratégies ? Quelles sont les conceptualisations sous-jacentes aux erreurs récurrentes ? Une autre limite tient à la difficulté à distinguer dans quelle mesure le degré d'accommodation est lié à un apprentissage ou à la difficulté propre d'un puzzle [5].

D'où la nécessité d'analyses plus approfondies pour relier étroitement modélisation du domaine et modélisation de l'apprenant, ce qui revient, dans le cadre d'analyse de la théorie des champs conceptuels de Vergnaud à déterminer de quelle manière un sujet construit un champ conceptuel des bases de la programmation en langage Scratch. Nous visons cet approfondissement dans la suite de nos travaux, notamment en menant des analyses à des niveaux de granularité plus fins : structure des programmes soumis et nature des erreurs commises, interactions avec l'interface (notamment mouvements de souris).

Remerciements

Ce travail est soutenu par les projets Interreg Teach Transition (https://teachtransition.eu) et ANR IE-CARE (https://iecare.lip6.fr/. Nous remercions aussi l'association France-ioi pour la mise à disposition de la plateforme chticode.algorea.org.

Références

1. Bouton, M., Peter, Y., Léonard, M., El Mawas, N. : Tentative de prédiction de la réussite à un exercice de programmation (2022)
2. Chevalier, M., Giang, C., Piatti, A., Mondada, F. : Fostering computational thinking through educational robotics : a model for creative computational problem solving. International Journal of STEM Education **7**(1), 1–18 (2020), publisher : SpringerOpen
3. Crahay, M. : Dangers, incertitudes et incomplétude de la logique de la compétence en éducation. Revue française de pédagogie. Recherches en éducation (154), 97–110 (2006), publisher : ENS Éditions
4. Jiang, B., Zhao, W., Zhang, N., Qiu, F. : Programming trajectories analytics in block-based programming language learning. Interactive Learning Environments **30**(1), 113–126 (2022), publisher : Taylor & Francis
5. Pelánek, R. : Bayesian knowledge tracing, logistic models, and beyond : an overview of learner modeling techniques. User Modeling and User-Adapted Interaction **27**(3), 313–350 (2017), publisher : Springer
6. Pelánek, R. : Exploring the utility of response times and wrong answers for adaptive learning. In : Proceedings of the fifth annual ACM conference on learning at scale. pp. 1–4 (2018)
7. Pelánek, R., Effenberger, T. : Design and analysis of microworlds and puzzles for block-based programming. Computer Science Education **32**(1), 66–104 (Jan 2022)
8. Piaget, J. : La naissance de l'intelligence chez l'enfant. Delachaux et Niestlé Neuchatel-Paris (1935)
9. Toll, D., Wingkvist, A. : Visualizing Programming Session Timelines. In : Proceedings of the 11th International Symposium on Visual Information Communication and Interaction. pp. 106–107 (2018)
10. Vergnaud, G. : Pourquoi parler autant de conceptualisation. In : Actes du colloque Conceptualisation et surdité Synthèse des travaux, Suresnes (2003)
11. Vergnaud, G. : La théorie des champs conceptuels. In : Recherches en didactique des mathématiques, vol. 10/2.3. La Pensée Sauvage (1991)
12. Vergnaud, G. : Au fond de l'action, la conceptualisation. Presses Universitaires de France (2011), pages : 275-292 Publication Title : Savoirs théoriques et savoirs d'action
13. Vergnaud, G. : Forme opératoire et forme prédicative de la connaissance. Investigações em Ensino de Ciências **17**(2), 287–304 (2012)
14. Vergnaud, G. : Qu'est-ce que la pensée ? La nouvelle revue de l'adaptation et de la scolarisation **N° 63**(3), 277–299 (2013), publisher : I.N.S.H.E.A.

La pédagogie par projet agile abat ses cartes : une métaphore ludique au service de la pratique

Alexis Lebis[0000−0003−2104−8671], Patrick Delaporte, Romain Deleau, Rémy Pinot et Mathieu Vermeulen[0000−0003−3646−1741]

IMT Nord Europe, Institut Mines-Télécom, Univ. Lille, Centre for Digital Systems
F-59000 Lille, France
lebis.research@pm.me

Résumé. La pédagogie par projet, et plus récemment la pédagogie par projet agile (P3A) est un paradigme d'apprentissage basé sur l'autonomisation et la socialisation des étudiants. Des difficultés d'opérationnalisation et d'engagement des apprenants (comportemental, motivationnel ou émotionnel) s'observent. Nous proposons un nouveau modèle de P3A qui s'inscrit dans la métaphore ludique du jeu de cartes pour conduire les projets. Ce modèle, implémenté dans un prototype, a été évalué avec des experts. Nous proposons deux nouveaux profils xAPI pour répondre au besoin de tracer les activités survenant dans la P3A afin de renforcer l'interopérabilité des outils de la communauté et le partage des traces.

Mots-clés : Pédagogie par projet, Méthodes agiles, xAPI, Jeu de carte

Abstract. Project-Based Learning, and more recently Agile Project-Based Learning, are rooted in a learning paradigm based on student empowerment and socialisation. Difficulties appear in operationalisation as well as in learner engagement (behavioural, motivational or emotional). We propose a new model of agile project-based learning that uses the playful metaphor of card games. It has been implemented in a prototype and tested by experts. Incidentally, we also propose two new standards for tracing the activities that occur in these learning situations to enhance the interoperability of community tools and the sharing of traces.

Keywords: Project-Based Learning, Agile Method, xAPI, Card Game

1 Introduction et contexte

La pédagogie par projet (P3) est aujourd'hui largement répandue dans l'enseignement supérieur. De nombreux travaux mentionnent son intérêt pour permettre aux étudiants d'acquérir des savoir-être et savoir-faire liés au domaine d'enseignement cible, mais aussi à la collaboration, à la construction des connaissances ou encore à la gestion de projet. Cette méthode pédagogique intègre l'évaluation des étudiants sous la forme d'un projet à mener, celui-ci devant permettre d'évaluer le niveau des participants. La P3 s'appuie souvent sur une vision classique de la réalisation du projet (*e.g.* cycle en V) [2]. Les approches agiles, telle SCRUM [10], sont parfois intégrés à cette pratique pédagogique [4]

avec un certain succès. L'intérêt principal de la pédagogie par projet agile (P3A) étant ainsi d'apporter un cadre formel à la réalisation du projet et professionnalisant.Toutefois la P3, et *a fortiori* la P3A, est un écosystème riche d'interactions complexes ancré dans un paradigme d'apprentissage par l'autonomisation et la socialisation des étudiants : sa bonne mise en oeuvre revêt de nombreuses difficultés pédagogiques. En effet, l'engagement comportemental, motivationnel ou émotionnel n'est pas toujours favoriser par la P3 [3]. Il semblerait également qu'elle ne soit pas adaptée à certains étudiants [6] en raison d'une trop grande liberté et d'une rupture forte par rapport aux méthodes pédagogiques plus classiques, mais aussi de la présence exacerbée de relations sociales [7]. De plus, les enseignants utilisant ce type d'approches n'évaluent souvent que les contributions relatives à leur expertises (*e.g.* un projet en informatique sera jugé par rapport à la qualité des livrables informatiques). Ainsi, le déroulement du projet n'est souvent que partiellement évalué, au même titre que l'organisation du groupe ou l'attribution réelle des tâches [5]. De fait, l'évaluation des compétences, savoir-faire et savoir-être liés à la gestion de projet pose question, notamment dans le cas la P3A qui nécessitent une organisation particulière. Une problématique émerge alors : *comment favoriser une démarche de P3A pour les acteurs impliqués, à l'aune de la réalité de la pratique, de la multi-dimensionnalité des interactions et des réalités individuelles (e.g. désengagement, dysfonctionnement du groupe, suivis de l'enseignant) qui y surviennent ?* Conséquemment, il semble pertinent de proposer un environnement adéquat qui donnerait un cadre organisationnel aux étudiants, ainsi qu'un moyen pour les enseignants de les suivre et de les d'accompagner dans leur apprentissage.

Cet article présente une contribution double. D'une part, nous présentons en section 2 le résultat de nos travaux proposant un modèle pouvant favoriser une démarche de P3A pour les apprenants et pour les enseignants. Ce modèle, qui exploite des mécaniques issues des jeux de carte, a été évalué avec des experts (cf. section 3). D'autre part, en section 4, nous proposons une spécification des traces que le modèle peut produire *via* de deux profils xAPI, afin de répondre aux besoins de production de données et d'analyse de la communauté. Elle certifie la qualité du processus de traçage et renforce l'interopérabilité des EIAH.

2 Proposition

Dans cette section, nous proposons un modèle ludifié pour favoriser une démarche de P3A pour les apprenants et pour les enseignants. La littérature attribuent à la ludification des effets bénéfiques sur les aspects cognitifs, comportementaux et motivationnels des étudiants [9] et ces effets semblent également s'observer lorsque appliquée à la P3 [1]. Or, dans le cadre d'une P3A, les étudiants évoluent intentionnellement dans un cadre formalisé (*e.g.* SCRUM). Il est donc important d'assurer que les transformations opérées sur ce cadre – ici une ludification de la pratique – n'interfèrent ni avec ledit cadre, ni avec les compétences transversales que les étudiants sont censés acquérir dans ce type de pratique, afin qu'elles soient exploitables en situation professionnelle. À l'aune des pratiques des différentes méthodes agiles et des outils associés (*e.g. kanban*,

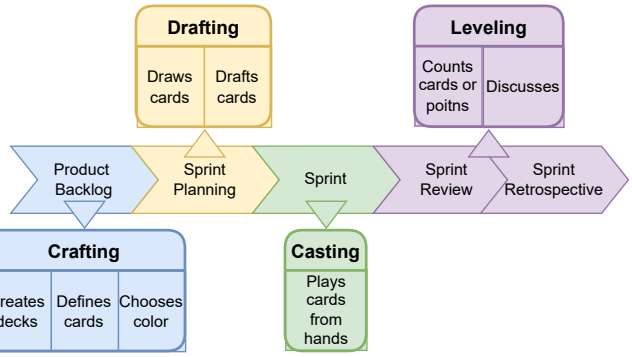


FIG. 1 – Schéma du *workflow* des méthodes agiles revisité dans une métaphore du jeu de cartes et ses 4 étapes impliquant des éléments ludiques.

poker planning), nous avons décidé d'inscrire la démarche de P3A dans une métaphore de jeux de cartes, bénéficiant ainsi des avantages de la ludification. Le modèle qui en résulte, illustré dans la Figure 1, permet de capturer le *workflow* et les éléments utilisés dans les méthodes agiles, notamment SCRUM : ils sont associés à divers éléments et pratiques de jeux de cartes classiques (*e.g.* pioche, défausse, *deck*). Nous avons également introduit la notion de traçabilité des activités (cf. section 4) afin d'appuyer de futurs mécanismes d'accompagnement et d'assistance pour les apprenants et les enseignants. Le modèle est décrit ci-après. Dans toutes les étapes du modèle, les enseignants peuvent intervenir.

Crafting Le *Product Backlog* est une organisation en amont du travail à effectuer pour le projet à réaliser. Il contient une liste ordonnée des exigences du produit que l'équipe doit accomplir, généralement sous la forme de *user stories* contenant des tâches. Dans notre métaphore, les apprenants sont alors amenés à créer des *decks* (*i.e.* paquets de cartes) en y définissant des cartes – analogie de la tâche. Chaque carte contient des informations (*e.g.* nom de la tâche, description) et peut se voir ajouter des modificateurs (*i.e.* des effets permanents), comme un nouveau type, ou des points. La carte appartient aussi à une famille – l'analogie en méthode agile serait la priorité de la tâche.

Drafting Le "*Sprint Planning*" correspond à la planification d'un *sprint* (*i.e.* l'intervalle de temps où l'équipe doit remplir les tâches assignées). C'est à ce moment que les différentes tâches sont sélectionnées et attribuées aux membres de l'équipe. Dans notre métaphore, les apprenants sont tout d'abord amenés à piocher un ensemble de cartes issus des différents *decks* pour créer un *booster* afin de préparer le *sprint*. Une fois les cartes piochées, une phase dite de *draft* commence avec tous les membres du groupe impliqués dans le projet : chaque joueur choisit une carte à tour de rôle du *booster* et la place dans sa "main de joueur". Il peut y avoir des contraintes appliquées sur cette main, comme un nombre maximal de carte ou une force de cartes maximale (*e.g.* analogie avec le nombre de jours/homme maximal pour un apprenant).

Casting Lors d'un *Sprint*, chaque membre de l'équipe tente de réaliser au fur et à mesure les objectifs et les tâches qui lui ont été assignées. Un *sprint* est toujours limité dans le temps. Dans notre métaphore, chaque apprenant possède une main de cartes. Il peut jouer une carte de sa main vers une zone de jeu pour indiquer qu'il pense avoir réalisé la tâche associée. La carte reste alors dans la zone de jeu le temps du *sprint*, mais peut être reprise pour indiquer que des modifications supplémentaires sont nécessaires pour la tâche. Il peut aussi avoir un coût associé à jouer une carte, pour l'analogie de la difficulté d'une tâche.

Leveling Lors d'un *Sprint Review* l'équipe présente le travail accompli, les tâches réalisées ou non et l'effet sur le projet, puis récupère différents *feedbacks*. Le *Sprint Retrospective* est une étape permettant de réfléchir sur le ou les *sprints* passés et d'étudier comment ils se sont déroulés, tant du point de vue humain que des interactions ou des outils. Dans notre métaphore, les apprenants se regroupent et, à tour de rôle, comptent leurs cartes jouées dans la zone de jeu – et éventuellement les points associés – puis les défaussent. Cette étape est une analogie de la validation des tâches (*i.e.* complétées et approuvées) d'un projet. Ils peuvent aussi s'exprimer sur la phase de jeu (*e.g.* faire une remarque, décerner une récompense) : ses interventions doivent alors être conservées et accessibles.

3 Expérimentation

Nous avons développé un prototype réifiant un sous-ensemble majeur du modèle présenté en section 2. Les quatre étapes et l'orchestration ont été implémentées ; seuls certains aspects ne l'ont pas été comme la notion de contraintes dans la main ou encore l'intervention à tout moment des enseignants dans le processus. Quelques images du prototype sont disponibles en ligne[11]. Nous avons conduit une expérimentation sous la forme de *focus group* avec des experts impliqués 1) dans l'enseignement (2 personnes), 2) dans le développement logiciel (3 personnes) et 3) le développement agile (1 personne). Cette expérimentation, d'une durée de 3 heures, avait pour but de mettre à l'épreuve la métaphore proposée et sa cohérence, notamment dans un cadre opérationnel et pédagogique (*e.g.* transfert de compétences). La première partie de l'expérimentation était destinée à la présentation du prototype durant 30 minutes. La deuxième partie consistait, pour chaque expert, à débuter la conduite d'un projet de leur choix. Ils étaient libre d'inscrire sur papier leurs remarques et leurs avis tout au long de l'expérimentation. Une phase de débriefing a finalement clos l'expérimentation.

Les retours expérimentaux sont encourageants quant à la transformation opérée par l'adoption de la métaphore, ce qui nous motive à l'utiliser pour des situations pédagogiques. Notre modèle semble convenablement permettre la conduite de projet : les experts ont réussi à mener à bien les objectifs et tâches qu'ils s'étaient fixés. De plus, ils ont unanimement pointé le caractère intéressant de notre proposition. Toutefois, une surcharge cognitive supplémentaire pour certains experts est à noter. Elle est imputable aux efforts pour faire correspondre à notre proposition des éléments issues de leur pratique de l'agilité et de développement, notamment le concept de tâche à réaliser qui se confond dans la

création et la distribution dans les mains des joueurs. Nous nous interrogeons si cette surcharge cognitive peut survenir dans le cas d'étudiants néophytes ou en cours de formation dans les méthodes agiles, et si cette surcharge s'opère dans l'autre sens : un expert de notre métaphore confronté à une méthode agile traditionnelle éprouvera-t-il les mêmes difficultés ? Cependant, ces retours sont également a mettre en perspectives de ceux concernant l'UX et le *Design* du prototype. Les experts pointent un manque de clarté de certaines interactions et de la navigation dans l'application, notamment lors de la phase de *Draft*. Nous documenterons les deux premières étapes lors des futurs expérimentations afin d'identifier s'il s'agit d'une surcharge liée à notre proposition ou notre prototype.

4 Tracer la conduite de projet & profiles xAPI

Nom	**Type xAPI**	**Description**
Deck	ActivityType	A stacked collection of Cards where cards are drawn from during a game.
Card	ActivityType	A card is a double sided rectangle with distinguishing motifs often interpreted as information. Normally the backs of the cards make them indistinguishable. The faces of the cards may all be unique, or there can be duplicates.
Hand	ActivityType	A cards collection belonging to the actor during a game, where card can be played. In general, all cards from a Hand are placed such that the face is in front of the actor, leaving the back to all the other players.
Pack	ActivityType	A pack represents a collection of stacked cards. It should be used as a broader term than other synonymous names used in card games such as "deck"
Discard Pile	ActivityType	A Pack which contains all the cards removed from the game. The cards are stacked either with their face or back visible. Usually, players can no longer play card in the Discard Pile. It is semantically equivalent to a Graveyard
Play zone	ActivityType	Virtual or physical space delimitation representing where cards can be played
Deals	Verb	Indicates that the actor has given one or several cards to others
Shuffles	Verb	Indicates that the actor has mixed cards into a different order before using them
Draws	Verb	Indicates that the actor pull the cards from the top of a Pack (usually a Deck). The drawn cards usually complete the Hand of the actor
Discards	Verb	Indicates that the actor consciously removes a card from the game. Usually cards are discarded either from the Hand or the Deck, and put in the Discard Pile
Plays	Verb	Indicates that the actor takes a card and puts it into the Play Zone. Usually, cards can only be played from the actor's Hand
Flips	Verb	The actor has changed the side of a card, thus hidding or revealing its information
Cost	Extension	How many a card cost to be played. Its meaning depends of the game
Value	Extension	Numerical value associated to a Card or a Collection
Content	Extension	Textual information written in the support, often explaining card's particularity
Title	Extension	Textual information helping to distinguish a specific card from all the others
Side	Extension	An information representing either the face of a card is visible or its back

TABLE 1 – Les éléments composant le profil xAPI associé aux jeux de cartes.

Le modèle proposé en section 2 a aussi été conçu pour répondre au besoin de traçage des activités survenant dans la P3 et P3A et, au vue des retombées pour la communauté, nous avons décidé d'élaborer un vocabulaire dédié en xAPI. Ce vocabulaire se décompose en deux profils xAPI (*i.e.* spécifications compréhensibles par l'homme et la machine) complémentaires destiné à la communauté pour capturer dans les différentes situations pédagogiques les activités associés aux démarches agiles ou aux jeux de cartes d'une manière générale. Bien que l'élaboration du profil *Agile* complètent des travaux sur les ontologies pour l'agile [8], notre profil *CardGame* constitue une contribution singulière : à notre connaissance, aucune ontologie dédiée aux jeux de cartes n'existe. Ainsi, grâce à ces profils, nous favorisons dans la communauté le partage et la réutilisation des traces produites et indicateurs issues de ces pratiques. Des applications pédagogiques exploitant des éléments de cartes (*e.g.* flashcard) ou agiles peuvent

dorénavant interopérer et entreposer leurs traces de manière homogène (*e.g.* LRS xAPI). En Table 1, par soucis de place, nous listons uniquement le profil xAPI dédié aux jeux de cartes. Les deux profils sont toutefois disponibles en ligne [11]. Nous avons opérationnalisé ces deux profils dans notre prototype avec succès.

5 Conclusion et perspectives

Dans ce papier, nous présentons un modèle ancré dans une métaphore ludique de jeux de cartes, destiné à favoriser une démarche de P3A à la fois pour les enseignants et pour les apprenants. Les premiers résultats expérimentaux obtenus auprès d'experts à partir d'un prototype confirme qu'il s'agit d'une piste intéressante à explorer pour une application pédagogique, avec toutefois une vigilance sur la surcharge cognitive liée à la transferabilité des compétences entre les méthodes agiles et la notre. Nous avons aussi présenté deux nouveaux profils xAPI qui offrent de nouvelles perspectives d'interopérabilités pour les EIAH.

Nous prévoyons maintenant de mener des expérimentations à l'échelle de la classe pour impliquer directement les apprenants et étudier les effets de nos propositions sur la réalisation de projets inscrits dans une démarche de P3A. En parallèle, nous prévoyons de tester notre outil pour la conduite de projets de recherche. Nous avons également prévu de transposer le profil xAPI pour les jeux de cartes dans une ontologie web structurée pour renforcer plus encore l'interopérabilité dans la communauté et les outils à disposition.

Références

1. Huang, W., Li, X., Shang, J. : Gamified project-based learning : A systematic review. In : Int. Conference on Blended Learning. pp. 313–324. Springer (2022)
2. John, W., Thomas, W. : A review of research on project-based learning. TA Foundation, California (2000)
3. Johnson, C.S., Delawsky, S. : Project-based learning and student engagement. Academic research international **4**(4), 560 (2013)
4. Laval, J., Fleury, A., Karami, A.B., Lebis, A., Lozenguez, G., Pinot, R., Vermeulen, M. : Toward an Innovative Educational Method to Train Students to Agile Approaches in Higher Education : The A.L.P.E.S. Education Sciences **11** (2021)
5. Lebis, A., Prior, E., Mandran, N., Karami, A., Vermeulen, M. : Promouvoir et soutenir la Pédagogie Par Projet Centré Humain dans le supérieur : le projet APACHES. In : DIDAPRO 8 - DIDASTIC. Lille, France (Feb 2020)
6. Leggett, G., Harrington, I. : The impact of project based learning (pbl) on students from low socio economic statuses : a review. International Journal of Inclusive Education **25**(11), 1270–1286 (2021)
7. Lobczowski, N.G., Lyons, K., Greene, J.A., McLaughlin, J.E. : Socioemotional regulation strategies in a project-based learning environment. Contemporary Educational Psychology **65**, 101968 (2021)
8. Ortega, W., Calvache, C., Pino, F. : Ontoagile : an ontology for agile software development processes. Dyna (Medellin, Colombia) **86**, 86–97 (03 2019)
9. Sailer, M., Homner, L. : The gamification of learning : A meta-analysis. Educational Psychology Review **32**(1), 77–112 (2020)
10. Schwaber, K., Beedle, M. : Agile software development with Scrum, vol. 1. Prentice Hall Upper Saddle River (2002)
11. Vermeulen, M., Lebis, A. : Apaches website – online resources for eiah2023 (2023), `https://apaches.wp.imt.fr/eiah23-deckify-rsc`

EIAH2023 : 11ème Conférence sur les Environnements Informatiques pour l'Apprentissage Humain
La conférence pluridisciplinaire francophone sur la conception et l'analyse des environnements numériques pour l'éducation et la formation
12-16 juin 2023 Brest (France)

Session de communications 3
Modélisation, exploitation et description de ressources pour la conception

Modélisation de la maturité numérique des enseignants

État de l'art et conception d'un modèle unifié : MUME

Christine Michel[1] et Laëtitia Pierrot[2]

[1] Université de Poitiers, UR TECHNE, Poitiers, France
christine.michel@univ-poitiers.fr
[2] Université de Franche-Comté, UR ELLIADD / UAR MSHE, Besançon, France
laetitia.pierrot@univ-fcomte.fr

Résumé. Cet article présente une revue de littérature des modèles de maturité numérique des enseignants. Dans ce cadre, nous avons extrait 11 modèles applicables au champ de l'enseignement scolaire obligatoire et proposons une synthèse des dimensions constitutives de chacun des modèles et de la manière dont ces dimensions contribuent à déterminer les niveaux de maturité numérique des enseignants. Si notre synthèse met en lumière la diversité des dimensions retenues dans les modèles, elle révèle aussi que la plupart de ces modèles ne fournissent qu'une image partielle de la maturité numérique. En outre, l'essentiel de ces modèles se concentre sur les derniers niveaux de maturité, associés à des enseignants innovateurs ou pionniers, et écarte les enseignants non utilisateurs ou faibles utilisateurs numériques, pourtant bien représentés dans le contexte français. La dernière partie de l'article consiste à proposer un modèle unifié de la maturité numérique des enseignants, MUME, répondant à ces deux problèmes.

Mots-clés: Modèle de maturité numérique, Intégration du numérique à l'école, Pratiques numériques des enseignants.

Abstract. This article presents a literature review of models for measuring teachers' digital maturity. From our review, we identified 11 models that are relevant to compulsory schooling. In this paper, we aim to synthesize the constituent dimensions of each model and explore how these dimensions contribute to determining the digital maturity levels of teachers. Our synthesis highlights the diversity of dimensions included in the models but also reveals that many of them provide only a partial view of technology maturity. Additionally, most of these models focus on the latest levels of maturity, associated with innovative or pioneering teachers, and do not consider non- or low digital user teachers, who are well represented in the French context. We propose a unified model of teachers' digital maturity to address these issues, called MUME.

Keywords: Technology maturity model, Technology integration in education, Teachers' digital practices.

1 Introduction

La transformation numérique est devenue l'un des enjeux les plus critiques dans le contexte éducatif [1]. En France, cette transformation est d'autant plus cruciale qu'une faible intégration des outils numériques s'observe dans les pratiques des enseignants du premier et du second degré[1]. De plus, le potentiel des technologies numériques pour l'enseignement et l'apprentissage ne dépend pas principalement du type de technologie ou de sa fréquence d'utilisation, mais plutôt de la façon dont les technologies numériques sont utilisées pour stimuler cognitivement les élèves et les engager dans des activités d'apprentissage [1]. Or, la crise sanitaire de 2020 a eu un effet de stimulation sur les usages numériques, mais ils restent encore limités à des pratiques de transmission de ressources et d'apprentissage passif [2]. Différents modèles, comme le TPACK [3], le SAMR [4], le NETS-T[5] ou le DigCompEdu [6] ont été proposés pour décrire les capacités, les dynamiques, les niveaux d'intégration ou de maturité numérique des enseignants, mais ces modèles sont relativement hétérogènes.

Dans son sens courant, la maturité désigne un état d'être complet, parfait ou prêt, qui s'inscrit dans un système [7]. Dans les contextes organisationnels, la maturité est l'objectif qui guide de nombreux besoins de transformations, c'est-à-dire des changements fondamentaux de stratégies, de structures et de répartition du pouvoir, et la transformation numérique peut être considérée comme un processus continu d'adoption par les collaborateurs d'une offre numérique en pleine mutation [7]. En éducation, les modèles de maturité s'intéressent aux différentes dimensions qui agissent sur l'intégration des technologies, en particulier le pilotage des actions de numérisation des structures et l'activité professionnelle des enseignants. Ainsi, au-delà des problématiques d'accès, de disponibilité et de fréquence d'usage, la maturité numérique considère-t-elle des questions de politique institutionnelle et de pédagogie (par exemple au niveau des contenus disciplinaires et pédagogiques traités, ou encore des changements attendus en ajoutant l'outil numérique) soulevées par l'introduction des technologies [8]. Les modèles de maturité trouvent aussi leur intérêt pour mesurer, diagnostiquer ou accompagner les enseignants dans leur utilisation du numérique [9]. Plus largement, l'étude de la maturité permet d'aborder l'adoption des outils numériques en combinant des facteurs liés à l'enseignant, et à son contexte d'exercice professionnel [10]. Pour cela, il est nécessaire de considérer l'apprenant, l'enseignant et leur contexte plus large en collectant des données pour mesurer l'étendue et la profondeur de l'intégration du numérique dans un établissement [11–13]. Cette approche socio-contextuelle de l'adoption du numérique se distingue des travaux portant sur le « bon » usage attendu du numérique, et invite à s'intéresser notamment aux enseignants, en tant « qu'agents de changement », voire de « leader » qui mettent en œuvre l'outil [5, 14]. L'état de l'art réalisé dans le cadre de cet article s'inscrit dans cette perspective.

[1] https://www.education.gouv.fr/media/95365/download

La multitude de modèles existants sur la maturité numérique des enseignants met en concurrence des observations empiriques de pratiques conceptualisées, des propositions théoriques non éprouvées sur le terrain et d'autres validés empiriquement. Or, ces modèles deviennent les socles à partir desquels se font des analyses empiriques, et des stratégies de formation des enseignants ou de diagnostics des établissements. Ces modèles sont aussi utiles pour construire des cursus de formation ou adapter les EIAH au profil des apprenants, qu'ils s'agissent d'élèves ou d'enseignants. En effet, tout comme le CRCN (adapté du DigComp) et Pix ont structuré les méthodes de certification et formation au numérique des élèves en France, les enseignants seront prochainement certifiés sur la base du CRCN-Edu (adaptation du DigCompEdu) et formés par une version spécifique de Pix : Pix+Edu[2].

L'objectif de notre étude est donc, à partir d'une revue de littérature, de faire une analyse des différents modèles de maturité et de proposer une version unifiée ayant une dimension holistique. Notre question générale de recherche (QR) est la suivante : quel modèle représente le mieux la maturité numérique des enseignants ? Spécifiquement, selon quels domaines la définir ? (QR1) Selon quels niveaux la caractériser ? (QR2).

2 Méthode

Nous avons travaillé selon une méthode de revue herméneutique [15], c'est-à-dire en identifiant : (1) les modèles de maturité des enseignants en partant des précédentes revues systématiques de littérature portant sur les modèles d'intégration et de maturité numérique en éducation [8–10, 14, 16, 17] et dans d'autres organisations [7, 18], (2) en suivant l'ensemble des travaux cités dans l'article ou citant l'article pour découvrir d'autres modèles jusqu'à ne plus en avoir de nouveaux. Seuls les modèles applicables au contexte de l'enseignement obligatoire ont été retenus, soit 21 modèles [19]. Nous avons ensuite comparé ces modèles en considérant [7, 9] : *le périmètre* (Générique G, ou Spécifique S), *la description de l'activité professionnelle* (Partielle P ou Globale G), *la place de l'apprenant* (Faible F ou Présente P), la *spécification de niveaux de maturité* (Oui O ou Non N), *l'utilité* (Accompagnement A, Descriptif De ou Diagnostic Di), *l'origine de conception du modèle* (Empirique E ou Théorique T) et *la validation* (Oui O ou Non N)). Sur cette base, nous avons choisi les 11 modèles qui sont : les plus génériques en termes de périmètre, de description de l'activité professionnelle, les plus précis en termes de description de niveaux et qui sont basés sur des études empiriques ou qui ont été validés.

2.1 Comparaison des modèles selon leurs caractéristiques de conception

La plupart des modèles (voir tableau 1) envisagent le contexte d'utilisation du numérique comme un élément générique, 2 modèles spécifient ce contexte. 5 modèles ont la particularité de vouloir considérer l'ensemble de l'activité professionnelle des enseignants, en incluant les tâches en dehors de la classe (préparation, planification, etc.).

[2] https://pix.fr/actualites/actualite-pix-edu/

L'activité d'enseignement est complétée, pour 6 modèles, par celle des apprenants. Les niveaux de maturité ne sont pas mesurés dans 4 modèles. Ces modèles ont principalement une utilité descriptive. Dans les 7 autres modèles, la maturité numérique des enseignants est considérée comme un élément de développement professionnel, d'où la présence d'outils de diagnostic voire de guides ou feuilles de route pour favoriser le déploiement des technologies.

La modélisation de l'intégration du numérique dans l'enseignement provient, pour l'essentiel, de travaux basés sur l'observation de pratiques : 5 d'entre eux ont un ancrage théorique précisé et 7 modèles ont fait l'objet d'une validation empirique.

Table 1. Synthèse des modèles en fonction de leurs principales caractéristiques

	Périmètre	Activité professionnelle	Place de l'apprenant	Niveau de maturité	Utilité	Origine du modèle	Validation
BECTA [20]	G	G	F	O	Di	E	N
CIT [14]	G	G	F	N	A	E	N
DigCompEdu [6]	S	G	P	O	A	T	O
ICAP [1, 21]	G	P	P	O	A	T	O
ICTE-MM [16]	G	G	P	O	A	E	O
LoTi [22, 23]	G	P	F	O	Di	T	O
NETS-T [5]	S	G	P	N	A	E	O
PICRAT [9]	G	P	P	N	De	T	N
SAMR [4]	G	P	F	O	De	E	N
TIM [24]	G	P	P	O	Di	E	O
TPACK [3, 25]	G	P	F	N	De	T	O

3 Les modèles de maturité

3.1 Modèles basés sur les dynamiques d'appropriation des enseignants

Selon Puentedura, SAMR encourage les éducateurs à « passer » des niveaux d'enseignement grâce à la technologie, tout en maintenant la valeur et l'importance de la pédagogie et le curriculum [4]. 4 étapes définissent le SAMR : *substitution, augmentation, modification et redéfinition* de la tâche d'enseignement avec la technologie. Le modèle a été développé à partir d'observations et sans fondements théoriques, il est pourtant largement utilisé et cité dans les travaux scientifiques [26].

Le CIT [14] considère le processus collectif de renforcement des connaissances d'un groupe (enseignants, chefs de bureau, directeurs) et la manière dont la culture de l'organisation peut soutenir (ou entraver) l'intégration des technologies éducatives dans les pratiques scolaires. Le modèle comporte 4 états, plutôt qu'étapes, pour signaler que ces états ne sont pas linéaires et peuvent être vécus simultanément par les enseignants : lorsqu'une nouvelle technologie est introduite, le collectif est d'abord dans une phase de *choc* qui précède une phase de *négociation* vis-à-vis de leurs préjugés sur l'objet (positive, s'ils jugent par exemple que la technologie peut leur faire

gagner du temps, négative à l'inverse), puis des phases d'*autonomisation* (durant lesquelles ils construisent les premiers usages) et *exploration* (durant lesquelles ils développent de nouveaux usages).

3.2 Modèle basé sur les dimensions de la maturité pour les enseignants

Le TPACK [3, 25] est considéré comme l'un des modèles les plus importants décrivant les compétences des enseignants pour un enseignement réussi avec la technologie. La valeur ajoutée du modèle est de ne pas considérer individuellement les connaissances technologiques (TK), du contenu (CK) et pédagogiques (PK), mais plutôt leurs interactions matérialisées par les zones de recouvrement (TCK, PCK, TPK). En 2019, TPACK évolue pour inclure les connaissances contextuelles (XK) sur la manière d'intégrer les contraintes organisationnelles et situationnelles [25]. Le succès des efforts des enseignants ne dépend ainsi uniquement de leur connaissance TK, PK, CK et de leurs chevauchements, mais aussi de leur capacité à les mettre en œuvre en fonction du contexte.

3.3 Modèles basés sur l'approche pédagogique retenue par les enseignants

ICAP [21] ne décrit pas spécifiquement le niveau de maturité ou la capacité des enseignants, mais plutôt les processus et niveaux d'engagement cognitifs (du plus au moins coûteux) des apprenants pour 4 types d'activités d'apprentissage : *interactives, constructives, actives et passives*.[1]. LoTi a pour objectif d'évaluer l'efficacité de la mise en œuvre du numérique à travers 7 niveaux (du niveau 0, pour la *non-utilisation*, au niveau 6, correspondant au niveau de *raffinement*. Conceptuellement, LoTi décrit 5 dimensions (*enseignement/apprentissage avec le numérique, évaluation avec le numérique, créativité des élèves, développement professionnel et citoyenneté numérique*). L'utilisation des outils et ressources numériques en classe pour l'enseignement et l'apprentissage est mesurée à l'aide d'outils validés empiriquement [22, 23] pour contribuer ensuite au développement professionnel des enseignants.

3.4 Modèles mixtes articulant l'efficacité pédagogique et les niveaux de maturité

PICRAT [9] considère le type d'apprentissage et l'engagement de l'apprenant avec l'outil (PIC, passif, interactif, créatif) et comment l'outil modifie la mise en œuvre pédagogique de l'activité (RAT, remplacement, amplification ou transformation de la pratique), soit 9 combinaisons possibles. Pour chacune des catégories, le modèle distingue les méthodes pédagogiques, les processus d'apprentissage des élèves et les objectifs didactiques. Le TIM [24] se matérialise sous la forme d'une matrice d'intégration technologique[3]. Il comporte 5 niveaux d'intégration technologique (entrée, adoption, adaptation, infusion et transformation) et 5 caractéristiques de l'environnement d'apprentis-

[3] http://mytechmatrix.org et https://fcit.usf.edu/matrix/matrix

sage (actif, collaboratif, constructif, authentique et orienté vers un objectif) qui s'articulent autour des meilleures pratiques. Il aide l'enseignant à choisir comment utiliser les outils technologiques pour atteindre les objectifs d'apprentissage.

3.5 Modèles mixtes articulant les compétences et les niveaux de maturité

Le DigCompEdu a été développé pour définir les compétences numériques des enseignants, pour tous les niveaux ou matières à enseigner, à l'échelle européenne [6]. DigCompEdu considère les compétences professionnelles, pédagogiques et de l'apprenant selon *6 domaines* (eux même décomposé en 3 à 6 sous-domaines) et *6 niveaux d'utilisation du numérique dans l'éducation*. Les NETS-T (Normes nationales de technologie éducative pour les enseignants) se décomposent en *5 domaines* décrits selon *4 types d'activités* [5]. Dans l'ensemble, ces normes sont conçues pour l'autodiagnostic et la création de programmes éducatifs permettant aux enseignants de changer les attitudes à l'égard des nouvelles technologies. Elles ont été élaborées grâce aux contributions d'acteurs divers de l'éducation [27].

3.6 Modèles descriptifs de la maturité des organisations

Le modèle développé par le Becta en 2008 est conçu pour aider les établissements d'enseignement supérieur à atteindre une maturité numérique à travers un outil d'autoévaluation autour de 5 domaines (*leadership, contexte, ressources, soutien à l'apprentissage et enseignement et apprentissage*) et *5 niveaux* à destination des décideurs et des enseignants [20]. Le modèle a été complété en 2018 [28] pour décrire les contextes et cultures scolaires favorisant le développement systématisé du numérique (l'intégration) par la gestion et le soutien aux activités d'enseignement et d'apprentissage.

ICTE-MM est une proposition qui a pour ambition de se rapprocher de standards internationaux [16] et s'inspire pour cela du modèle standardisé du CMMI[4] (Capability Maturity Model Integration) et des travaux de l'ISTE [5]. Le modèle distingue 3 dimensions susceptibles de soutenir les processus éducatifs (*pilotage, stimulation de la culture numérique, ressources d'information et TIC*). ICTE-MM propose un outil d'autoévaluation et une feuille de route pour guider les chefs d'établissement sur la gestion du numérique.

4 Vers un modèle unifié de la maturité numérique des enseignants : MUME

4.1 MUME : les domaines descriptifs

Les domaines de caractérisation des modèles ont été structurés en considérant les modèles d'intégration les plus généraux pour aller vers les plus spécifiques. Les modèles ont été intégrés de manière à préserver au maximum les domaines et la structure de

[4] https://www.cmmiinstitute.com/

chaque modèle. Différentes restructurations ont été opérées de manière à articuler les modèles entre eux dans une vue unifiée (voir figure 1). Lorsqu'un domaine était déjà présent, il n'a pas été affiché dans la structuration. Ainsi l'ensemble des domaines du LoTi n'apparaissent pas explicitement dans la modélisation, mais ils sont considérés dans la mesure où ils sont déjà présents dans les autres modèles. 4 modèles structurent cette vue unifiée: le TPACK, le ICTE-MM, le DigCompEdu et l'ICAP. Nous avons choisi de ne faire apparaitre dans le TPACK que les dimensions qui concernent l'intégration des technologies, soit TPCK et XK.

Nous avons opéré des restructurations sur l'ICTE-MM et le DigCompEdu : les élèves ont été intégrés aux domaines du DigCompEdu qui concernent l'enseignant puisque ce sont les actions de l'enseignant vers les élèves qui sont considérés et non les actions des élèves eux-mêmes. Ainsi ces domaines sont-ils rattachés sous l'enseignant, dans la gestion des élèves. D'autres sous-domaines (du domaine 3) du DigCompEdu ont été restructurés autour : des pratiques pédagogiques de l'ICAP (pour intégrer le TIM et le PICRAT) et de la gestion des élèves (pour intégrer le sous-domaine « conseil »). La gestion de l'éducation de l'ICTE-MM a été, de la même manière, intégrée au domaine des administrateurs. Les dimensions du BECTA ont pu être ajoutées sur cette base. Les domaines du NETS-T (en jaune, à droite, dans la figure 1), ne sont pas cohérents avec les autres, car structurés par rôle plutôt que par compétence, mais ont été ajoutés pour faciliter, dans une perspective UX design, la conception de moyens ou de services d'accompagnement à la montée en maturité.

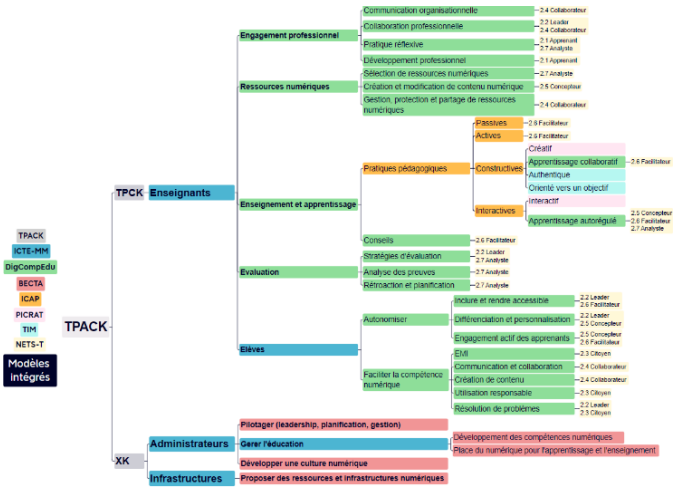


Fig. 1. (a) Modèles considérés (b) Critères de caractérisation des modèles de maturité

Le modèle unifié proposé comporte 3 domaines principaux (issus de l'ICTE-MM) : enseignants, administrateurs et infrastructure. Le domaine enseignant a été réduit à 4 sous-domaines (provenant du DigCompEdu) : engagement professionnel, ressources numériques, enseignement et apprentissage, évaluation et élèves. Les sous-domaines « enseignement et apprentissage » et « évaluation » pourraient être fusionnés, comme

dans le CRCN-Edu, mais la spécificité des sous-domaines de « évaluation » qui correspond à un rôle d'analyste doit être distinguée.

4.2 MUME : les niveaux d'intégration

Lorsque l'on compare les différents modèles en termes de niveau (voir tableau 2), on perçoit de nombreuses différences. Rares sont les modèles à ne pas mentionner de gradation dans les pratiques ou compétences (comme le TPACK ou les NETS-T). Les autres modèles considèrent un nombre de niveaux variant de 3 à 7. Seuls le DigCompEdu, le NETS-T et les modèles de maturité organisationnelle considèrent le rôle de leader que peut jouer l'enseignant dans la diffusion des usages et pratiques par la collaboration et le partage. Cette activité étant critique pour la diffusion des usages, nous choisissons de la conserver. De la même manière, seuls le CIT, le TPACK, le LoTi et les modèles de maturité organisationnelle considèrent la non-utilisation. Dans la mesure où toutes les pratiques ne sont pas instrumentées et que le choix de ne pas instrumenter ses pratiques ne relèvent pas nécessairement d'un manque de compétence chez les enseignants (hors contexte COVID), mais plutôt d'un choix pédagogique, nous conservons cette catégorie et intégrons dans la population un groupe de non-utilisateurs.

Les modèles ont aussi différents points communs. À part le NET-S, tous intègrent une gradation de la maturité allant d'un niveau « entrée », qui correspond aux usages les plus simples, à un niveau « transformation », correspondant à la création d'innovation d'usage avec la technologie. Dans la plupart des modèles (SAMR, CIT, ICAP, LoTi, TIM, DigCompEdu, BECTA, ICTE-MM) cette gradation considère l'expertise en termes de compétences techno-pédagogiques avec un cœur à 4 niveaux, globalement aligné sur les définitions de l'ICAP (passif, actif, collaboratif, interactif) et un niveau 5 qui correspond à la capacité à innover vers de nouvelles formes techno-pédagogiques.

Si on ajoute un niveau de « non-utilisation » au modèle de diffusion de l'innovation (ou *DOI*) de Rogers [29], on peut constater que les niveaux 6, 5, 3, 2, 1 (« Innovateur », « Adoptant précoce », « Retardataire » et « Non-utilisateur ») sont cohérents pour l'ensemble des modèles. Les niveaux 1 et 2 correspondent respectivement à une non-maturité et une entrée dans le processus d'intégration des technologies principalement par des pratiques de conception simple et de transmission de supports de formation. Le niveau 3 (majorité active) correspond à une phase d'exploration des possibilités et se concrétise au niveau 4 (et majorité précoce) par des stratégies pédagogiques actives. Le niveau 5 est plutôt caractérisé par des pratiques de leadership et de partage vers les autres membres de la communauté, ainsi que de gestion et d'analyse. Le niveau 6 est caractérisé par des capacités d'innovation et de maitrise complète de l'intégration des technologies. Le niveau 4 est moins cohérent. Il est souvent distingué dans les modèles spécifiques pour l'éducation en 2 niveaux : (expert, intégrateur) pour le DigCompEdu, (Infusion, Intégration) pour le LoTi, (Infusion, Adaptation) pour le TIM. Cette distinction ne nous semble utile que pour identifier les pratiques interactives et collaboratives. En effet, les pratiques interactives sont actuellement peu développées et pourraient correspondre à un niveau « adoptant précoce », mais ne correspondent pas à la capacité de diffusion et leadership de cette catégorie. Nous choisissons donc dans un premier temps

de les intégrer au niveau 4 et vérifierons la cohérence de ce choix par une étude empirique.

Table 2. Synthèse des modèles en fonction des niveaux de maturité

Modèles	Description des niveaux							Nb niveaux
DOI	Innovateur	Adoptant précoce	Majorité précoce		Majorité acquise	Retardataire		5
ICTE-MM	5 Optimisé.	4 Géré	3 Défini			2 Dévelop-pement	1 Initial	5
BECTA	5. Maturité	4. Avancé	3. Compétent		2. Autonomisation		1. Pas de maturité	5
DigCompEdu	Pionnier (C2)	Leader (C1)	Expert (B2)	Intégrateur (B1)	Explora-teur (A2)	Nouvel arri-vant (A1)		6
LoTi	Perfection-nement	Expansion	Intégra-tion	Infusion	Explora-tion	Sensibilisa-tion	Non-utili-sation	7
ICAP			Interac-tif	Collabora-tif	Actif	Passif		4
PICRAT	Transformation		Amplification			Remplace-ment		3
TIM	Transfor-mation		Infu-sion	Adaptation	Adoption	Entrée		5
SAMR	Redéfini-tion	Modification			Augmenta-tion	Substitution		4
CIT Model	Explorateur	Autonomisation				Négociation	Choc	4
TPACK	TPCK						TK, PK, CK	1
NETS-T								Pas de niveaux
Synthèse	**Transfor-mation**	**Développe-ment**	**Intégration**		**Amélora-tion**	**Substitution**	**Non-utili-sation**	**6**
	Pionnier	**Leader**	**Expert**		**Explora-teur**	**Nouvel arrivant**	**Non-utili-sateur**	**6**

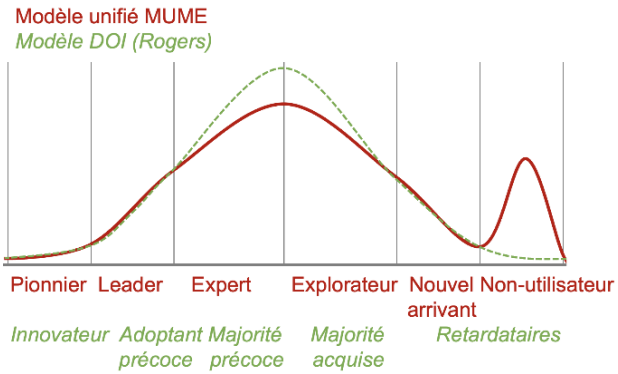


Fig. 2. Comparaison des courbes de diffusion des technologies pour l'éducation pour le DOI de Rogers [29] et le modèle unifié MUME

Nous recommandons d'utiliser un modèle en 6 niveaux qui considère soit les processus caractéristiques (*Transformation, Développement, Intégration, Amélioration, Substitu-*

tion, Non-utilisation), soit les pratiques des acteurs (*Pionnier, Leader, Expert, Explorateur, Nouvel arrivant, Non-utilisateur*). La courbe correspondante est représentée dans la figure suivante (voir figure 2), à titre indicatif. Des études empiriques doivent être menées pour en définir la forme précise. La classification de Rogers [29] y a aussi été présentée à titre de comparaison.

5 Discussion et perspectives

Cet article propose une revue de littérature sur la maturité numérique des enseignants. Nous avons identifié 11 modèles : 9 modèles spécifiques aux pratiques professionnelles des enseignants (sections 2.1 à 2.5) et 2 modèles portant sur leur contexte professionnel (section 2.6). Sur la base d'une analyse comparée de ces modèles, nous proposons un modèle unifié MUME, considérant les aspects individuels liés à l'enseignant, et les aspects organisationnels et contextuels. Ce choix permet de l'utiliser pour des travaux globaux sur l'intégration du numérique en éducation ou pour d'autres considérants uniquement l'enseignant. En outre, ce modèle présente l'avantage de couvrir l'ensemble de l'activité professionnelle de l'enseignant, plutôt qu'uniquement ses tâches d'enseignement. Le modèle unifié est composé de 6 niveaux cohérents globalement avec ceux du DigCompEdu, du DOI de Rogers et du ICTE-MM. Il a, en outre, la particularité d'intégrer un niveau de maturité 0 (niveau 1), correspondant à une non-utilisation que nous considérons comme un choix de l'enseignant plutôt qu'un frein, et de fusionner les niveaux B1 et B2 du DigCompEdu. Ce choix se justifie par le fait de proposer un outil mobilisable, à terme, pour du diagnostic et de l'accompagnement à l'intégration du numérique.

Notre modèle unifié de maturité constitue une première contribution à l'observation et l'analyse des niveaux de maturité numérique des enseignants. À ce stade, nous identifions deux perspectives pour ce travail. La première va consister à aller sur le terrain pour confirmer (ou rejeter) les propositions effectuées. À ce sujet, les projets de déploiement du numérique à l'école, tels que ceux développés dans le cadre des Territoires numériques éducatifs[5], représentent une bonne opportunité. En effet, ces projets, à l'échelle de départements français, envisagent le numérique comme facteur de transformation systémique. La prise en compte du contexte et de ses spécificités est, dans ce cadre, indispensable et nous pensons que notre modèle unifié répond à cet enjeu. La seconde perspective a trait aux modes de collecte des données pour mesurer les niveaux de maturité. Nous allons étudier : les outils existants et associés à certains modèles basés sur des questionnaires (le SELFIE pour le DigCompEdu, le TPACK-TS pour le TPACK, etc.), et les perspectives de collecte automatique de données. En effet, si ces outils sont parfois validés empiriquement, ils s'appuient principalement sur de l'autoévaluation, en dépit de la recommandation d'inclure d'autres modes de collecte de données [7, 30]. Notre objectif est, à terme, de s'appuyer sur les potentialités offertes par les travaux des Learning et Teaching Analytics pour proposer une approche mixte d'observation et d'analyse des niveaux de maturité numérique des enseignants.

[5] https://tne.reseau-canope.fr/

Remerciements . Ce travail a été réalisé en collaboration avec l'entreprise *Open Digital Education* et financé dans le cadre du projet *CoAI-DataStim* (Académie de Paris), du projet *TNE25* (Région académique Bourgogne-Franche-Comté) et du programme *NEXT* (Maison des sciences de l'homme et de l'environnement).

References

1. Antonietti, C., Schmitz, M.-L., Consoli, T., Cattaneo, A., Gonon, P., Petko, D.: "Development and validation of the ICAP Technology Scale to measure how teachers integrate technology into learning activities." Comput. Educ. 192, (2023).
2. Michel, C., Pierrot, L.: Towards Modelling the Technology Integration in Elementary School. A Diachronic Study of Teachers' Digital Practices During and After Covid-19 Lockdown. In: Hilliger, I., Muñoz-Merino, P.J., De Laet, T., Ortega-Arranz, A., and Farrell, T. (eds.) Educating for a New Future: Making Sense of Technology-Enhanced Learning Adoption. pp. 201–214. Springer International Publishing, Cham (2022). https://doi.org/10.1007/978-3-031-16290-9_15.
3. Mishra, P., Koehler, M.J.: Technological Pedagogical Content Knowledge: A Framework for Teacher Knowledge. Teach. Coll. Rec. 108, 1017–1054 (2006).
4. Puentedura, R.: SAMR - A research perspective. (2012).
5. ISTE: ISTE Standards for Educators, https://www.iste.org/standards/iste-standards-for-teachers, last accessed 2023/01/09.
6. Redecker, C.: European Framework for the Digital Competence of Educators: DigCompEdu. Punie, Y. (ed). EUR 28775 EN. Publications Office of the European Union, Luxembourg (2017).
7. Teichert, R.: Digital Transformation Maturity: A Systematic Review of Literature. Acta Univ. Agric. Silvic. Mendel. Brun. 67, 1673–1687 (2019).
8. Franklin, C., Bolick, C.: Technology Integration: A Review of the Literature. Presented at the Society for Information Technology & Teacher Education International Conference March 26 (2007).
9. Kimmons, R., Graham, C.R., West, R.E.: The PICRAT Model for Technology Integration in Teacher Preparation. Contemp. Issues Technol. Teach. Educ. 20, 176–198 (2020).
10. Harrison, C., Tomás, C., Crook, C.: An e-maturity analysis explains intention–behavior disjunctions in technology adoption in UK schools. Comput. Hum. Behav. 34, 345–351 (2014).
11. Underwood, J., Baguley, T., Banyard, P., Coyne, E., Farrington-Flint, L., Selwood, I.: Impact 2007: Personalising Learning with Technology, http://webarchive.nationalarchives.gov.uk/20110130111510/http://research.becta.org.uk/index.php?section=rh&catcode=_re_rp_02&rid=14202, last accessed 2021/04/30.
12. Underwood, J., Dillon, G.: Capturing complexity through maturity modelling. Technol. Pedagogy Educ. 13, 213–225 (2004).
13. Underwood, J., Baguley, T., Banyard, P., Dillon, G., Farrington-Flint, L.: Understanding the Impact of Technology: Learner and School level factors 2010. (2010).

14. Leite, L., Lagstedt, A.: The Collective Integration of Technology (CIT) Model: Helping Teachers Incorporate Technology Meaningfully in their Everyday Work. Int. J. Educ. Dev. Using Inf. Commun. Technol. 17, 249–268 (2021).
15. Sackstein, S., Matthee, M., Weilbach, L.: Theories and Models Employed to Understand the Use of Technology in Education: A Hermeneutic Literature Review. Educ. Inf. Technol. (2022).
16. Solar, M., Sabattin, J., Parada, V.: A Maturity Model for Assessing the use of ICT in School Education. Educ. Technol. Soc. 16, 206–218 (2013).
17. Carvalho, J., Pereira, R.H., Rocha, Á.: Maturity models of education information systems and technologies: A systematic literature review. (2018).
18. Pee, L.G., Kankanhalli, A.: A Model of Organisational Knowledge Management Maturity Based on People, Process, and Technology. J. Inf. Knowl. Manag. 08, 79–99 (2009).
19. Michel, C., Pierrot, L.: Les modèles de la maturité numérique des enseignants Lot 1 - État de l'art. (2023).
20. BECTA: Measuring e-maturity amongst work-based learning providers 2008 : final report. British Educational Communications and Technology Agency (BECTA) (2008).
21. Chi, M.T.H., Adams, J., Bogusch, E.B., Bruchok, C., Kang, S., Lancaster, M., Levy, R., Li, N., McEldoon, K.L., Stump, G.S., Wylie, R., Xu, D., Yaghmourian, D.L.: Translating the ICAP Theory of Cognitive Engagement Into Practice. Cogn. Sci. 42, 1777–1832 (2018).
22. Moersch, C.: Levels of Technology Implementation (LoTi): A framework for measuringclassroom technology use. Learn. Lead. Technol. 23, (1995).
23. Stoltzfus, J.: Determining Educational Technology and Instructional Learning Skill Sets (DETAILS): A New Approach to the LoTi Framework for the 21 st Century. (2006).
24. Kozdras, D., Welsh, J.: Enter the Matrix: A Pedagogy for Infusing Technology. In: Society for Information Technology & Teacher Education International Conference. pp. 536–541. Association for the Advancement of Computing in Education (AACE) (2018).
25. Mishra, P.: Considering Contextual Knowledge: The TPACK Diagram Gets an Upgrade. J. Digit. Learn. Teach. Educ. 35, 76–78 (2019).
26. Blundell, C.N., Mukherjee, M., Nykvist, S.: A scoping review of the application of the SAMR model in research. Comput. Educ. Open. 3, (2022).
27. Crompton, H., Sykora, C.: Developing instructional technology standards for educators: A design-based research study. Comput. Educ. Open. 2, (2021).
28. Ristić, M.: E-Maturity in Schools. Croat. J. Educ. 19, 317–334 (2018).
29. Rogers, E.M.: Diffusion of Innovations. Free Press, USA (2003).
30. Tomczyk, Ł., Fedeli, L.: Digital Literacy among Teachers -Mapping Theoretical Frameworks: TPACK, DigCompEdu, UNESCO, NETS-T, DigiLit Leicester. Presented at the November 1 (2021).

Construction et exploitation d'un référentiel de types de tâches d'apprentissage de la programmation

Sébastien Jolivet[1][0000−0003−3915−8465], Eva Dechaux[2][0000−0001−8579−1398], Anne-Claire Gobard[3][0000−0002−2866−084X] et Patrick Wang[4][0000−0003−3117−8189]

[1] IUFE & TECFA, Université de Genève, Suisse
sebastien.jolivet@unige.ch
[2] IUFE, Université de Genève, Suisse
eva.dechaux@unige.ch
[3] Lycée Kastler, Académie de Versailles, France
anne-clair.gobard@ac-versailles.fr
[4] Haute école pédagogique du canton de Vaud, Lausanne, Suisse
patrick.wang@hepl.ch

Résumé. Dans cette contribution, nous décrivons la construction d'un référentiel de types de tâches pour l'apprentissage de concepts fondamentaux de la programmation à l'aide d'une analyse systématique de moyens d'enseignement français et suisses. Cette analyse a permis d'identifier et structurer une centaine de types de tâches portant sur l'acte de programmer ainsi que sur les concepts de variables, instructions conditionnelles et boucles. Nous avons ensuite exploité ce référentiel pour analyser trois EIAH pour l'apprentissage de la programmation avec Python.

Mots-clés : Apprentissage programmation , modélisation du savoir , analyse d'EIAH

Abstract. This paper presents the construction of a repository of types of tasks that students might have to perform while learning the fundamentals of programming. To do so, we analyzed French and Swiss teaching and learning resources and identified around 100 types of tasks relating to the act of programming, the concepts of variables, conditional statements, and loops. We then used this repository to analyze three web-based tools designed to learn Python programming.

Keywords: Programming , Knowledge modelling , Tool Analysis

1 Introduction

L'apprentissage de la programmation est apparu dans les programmes scolaires du secondaire depuis maintenant quelques années aussi bien en France (SNT puis

NSI) qu'en Suisse romande (informatique en tant que discipline obligatoire). Les curriculums de ces disciplines mentionnent l'apprentissage de concepts fondamentaux de la programmation impérative. Si les grandes lignes sont définies en termes de notions à rencontrer, il n'y a pas ou peu d'éléments disponibles sur ce que signifie vraiment "apprendre à programmer" : quelles tâches doivent être rencontrées ? quelles procédures permettant de les réaliser doivent être enseignées ? basées sur quels savoirs ? Poser ces questions revient à interroger la transposition didactique externe [7] des savoirs de référence en informatique. Un des moyens d'étudier l'état de la transposition didactique est d'analyser des moyens d'enseignement de manière à identifier tâches, procédures et savoirs présents dans les premiers apprentissages de la programmation. Cet existant étudié, nous l'avons structuré (Sect. 3), puis nous avons questionné sa prise en compte dans trois EIAH d'apprentissage de la programmation avec le langage Python (Sect. 4).

2 Problématique et méthode

Décrire les savoirs qui doivent être travaillés pour l'apprentissage de la programmation n'est pas un problème nouveau. Par exemple, deux référentiels ont été produits, autour de la programmation récursive et de la programmation fonctionnelle, en lien avec le projet Comper [14]. Couderette [9] a, elle, étudié les effets de l'introduction de l'algorithmique dans le programme de mathématiques de seconde et a identifié des praxéologies algorithmiques ou informatiques pour décrire les tâches données aux élèves par des enseignants. Strock et Artaud [15] se sont aussi intéressé à l'apprentissage de l'algorithmique mais en étudiant un curriculum. Enfin, Chaachoua et al. [6] ont examiné l'articulation entre des tâches de programmation et des tâches d'apprentissage de la division euclidienne.

À notre connaissance, l'analyse de la transposition didactique liée à l'introduction de l'informatique comme discipline scolaire, réalisée en étudiant divers moyens d'enseignement n'a pas été faite. Ce constat, couplé au développement d'EIAH d'apprentissage de la programmation, notamment de Python, nous a amenés à formuler les deux questions de recherche suivantes :

QR1 Quelles tâches sont proposées dans les ressources d'enseignement, dans le cadre des premiers apprentissages d'un langage impératif ?

QR2 Quelles tâches sont présentes dans les EIAH d'apprentissage de Python ?

Pour décrire et structurer les tâches des moyens d'enseignement nous utilisons le modèle praxéologique issu de la théorie anthropologique du didactique [8]. Nous présentons maintenant ce modèle et notre méthode pour répondre à (QR1).

2.1 Modèle praxéologique

Le modèle praxéologique [8] postule que toute activité humaine peut être représentée à l'aide d'une praxéologie qui est définie par un quadruplet $[T, \tau, \theta, \Theta]$. Dans ce quadruplet, T est un *type de tâches*, les tâches le constituant pouvant être réalisées à l'aide de la *technique* τ. Le bloc $[\theta, \Theta]$ constitue le bloc du *logos*

et contient les éléments (définitions, propriétés, théorèmes du domaine) qui permettent de valider, justifier, expliquer le fonctionnement de la technique τ. Dans l'extension du modèle praxéologique T4-TEL [5], les techniques sont décrites à l'aide de types de tâches : nous parlons des *ingrédients de la technique*.

Par exemple si l'on considère le type de tâches *Déterminer le nombre d'itérations d'une boucle bornée donnée* (T_1), alors une technique τ_1 permettant de réaliser T_1 peut être décrite (partiellement) à l'aide des types de tâches *Identifier la structure de données parcourue par l'itérateur* et *Déterminer les valeurs prises par l'itérateur*. Une autre technique peut être décrite à l'aide du type de tâches *Exécuter pas à pas le programme et compter le nombre d'itérations*. Chacun de ces types de tâches admet sa propre technique, et le bloc du *logos* comprend les définitions de boucle bornée et d'itération, les propriétés de l'itérateur, etc.

2.2 Méthode d'analyse des moyens d'enseignement

Par programmation nous entendons, comme proposé par exemple dans [15], le travail qui consiste à produire un programme informatique permettant de réaliser une tâche t (trier une liste, déplacer un personnage, construire une figure avec une tortue, etc). Ce travail comprend le travail algorithmique et le travail d'implémentation. Nous limitons notre analyse de la manière suivante :

- Nous nous centrons sur les premiers apprentissages et plus précisément les types de tâches élémentaires de l'activité de programmation et ceux qui concernent les variables, les boucles et les instructions conditionnelles.
- Nous ne nous intéressons pas aux tâches liées à l'utilisation d'un langage spécifique (par exemple l'indentation dans le langage Python).

Ces limites fixées, nous avons procédé à l'analyse systématique, outre les curriculums, d'un corpus de ressources d'enseignement composé de ressources de natures différentes (notamment, manuels scolaires [2,4,11], cahier d'exercices [10], site Web contenant des moyens d'enseignement de l'informatique [1]) provenant d'institutions différentes (françaises et suisses et de niveaux scolaires différents). Nous avons listé l'ensemble des éléments de praxéologies identifiés dans ce corpus, il s'agit essentiellement des types de tâches et de quelques éléments relatifs aux techniques ou au bloc du *logos*.

Il est notable, dans le corpus étudié, que les techniques ne sont pas présentées de manière explicite. Par exemple on ne trouve pas la réponse à une question du type "comment fait-on pour concevoir une boucle bornée ?", celle-ci étant quasi exclusivement traitée au moyen d'exercices. Certains ingrédients de techniques font l'objet d'exercices spécifiques, mais le travail permettant à l'élève de "recoller les morceaux" n'est pas pris en charge dans les documents étudiés.

Les éléments du bloc du *logos* sont largement absents des ressources du corpus. Et, même si l'on rencontre parfois une définition, l'articulation entre les éléments du bloc du *logos* et les techniques est totalement absente. On ne trouve par exemple pas de justification du type "je peux utiliser telle ou telle structure comme itérateur dans une boucle bornée car...".

Dans la suite de cet article, nous nous concentrons sur le bloc praxique $[T, \tau]$. Le travail d'analyse systématique réalisé nous a permis d'obtenir une liste non or-

ganisée d'environ deux cents types de tâches. Un premier travail d'identification des types de tâches identiques, modulo des différences de vocabulaire, a produit une liste réduite d'une centaine de types de tâches. Le travail de structuration de ces types de tâches est l'objet de la section suivante.

3 Types de tâches d'apprentissage de la programmation

La structuration globale, que nous motivons dans les lignes suivantes, est présentée en Fig. 1. Tout d'abord nous considérons, comme évoqué précédemment, que programmer c'est produire un programme P_t qui réalise une tâche t. Nous identifions donc une première catégorie de types de tâches que nous nommons *Produire*. Or, pour produire un programme, en adoptant une posture anthropologique, nous identifions deux situations : 1) nous disposons d'un programme $P_{t'}$ qui réalise déjà une tâche t' proche de t ; 2) nous ne disposons pas d'un tel programme. Ceci nous amène donc à distinguer deux situations que nous avons nommées *Code existant* et *Feuille blanche*. Que l'on soit dans le cas initial de la

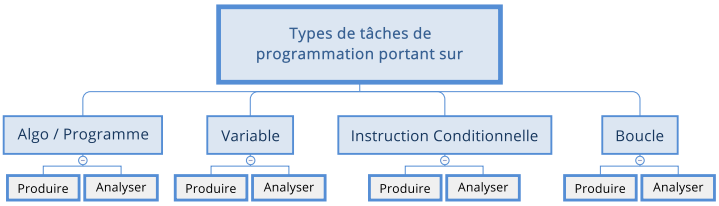


Fig. 1. Structuration globale des types de tâches.

feuille blanche, avec un premier code produit, ou dans le cas d'un code existant, nous avons, à un moment, un programme $P_{t'}$ qui ne réalise pas t. Il faut donc analyser $P_{t'}$ pour le modifier jusqu'à obtenir P_t. Ceci nous amène à définir une seconde catégorie de types de tâches que nous nommons *Analyser*. D'une manière générale, ces types de tâches sont des ingrédients de techniques des types de tâches de la catégorie *Produire*. Ils peuvent cependant avoir une existence autonome dans des exercices présents dans des ressources d'apprentissage. Ces types de tâches relatifs à l'objet P_t constituent la catégorie *Programme*.

Dans un deuxième temps, pour pouvoir produire P_t, ou travailler dessus, il faut mobiliser des objets et structures élémentaires. Nous en considérons trois : la première, *Variable*, concerne l'objet fondamental que sont les variables ; les deux autres, *Instruction conditionnelle* et *Boucle*, concernent les deux structures de contrôles incontournables. Au sein de chacun de ces ensembles, les types de tâches sont organisés selon les deux activités *Produire* et *Analyser*.

Une fois cette structuration définie, nous avons organisé les différents types de tâches identifiés au sein de chaque catégorie. Les différents types de tâches de la catégorie *Programme* sont présentés en Fig. 2.

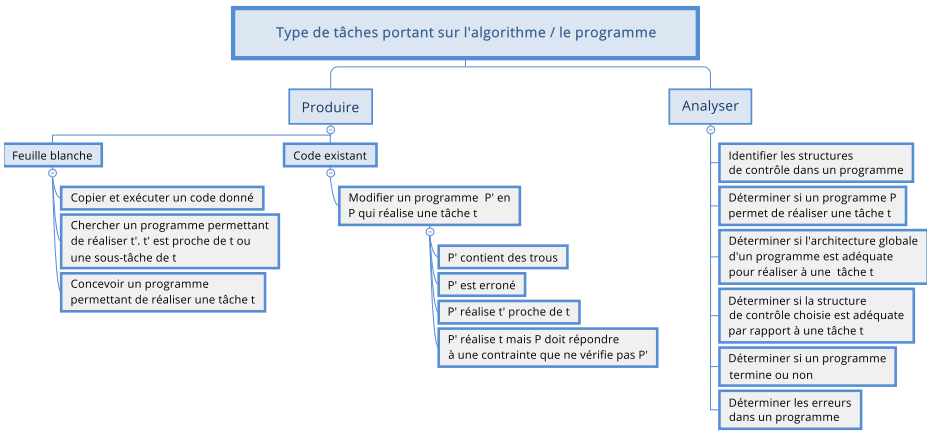


Fig. 2. Les types de tâches de la catégorie *Programme*.

Deux éléments sont importants à noter. Premièrement, il ne s'agit pas d'une liste exhaustive mais de types de tâches existants dans le corpus étudié. Deuxièmement, les types de tâches proposés sont de niveaux de granularité différents. Ainsi, le type de tâches *Concevoir un programme permettant de réaliser une tâche t* (noté T_P) est le plus global, tandis que le type de tâches *Déterminer une structure de contrôle pertinente par rapport au problème* est un ingrédient d'une technique permettant de réaliser T_P. Pour des raisons de place et de lisibilité nous ne présentons pas dans les arbres l'intégralité des types de tâches identifiés. Nous avons conservé tous ceux qui permettent de mener à bien l'analyse présentée dans la Sect. 4, mais avons supprimé des types de tâches ingrédients de technique comme par exemple *Identifier les variables nécessaires pour concevoir un programme* qui fait l'objet d'exercices spécifiques dans le corpus mais est en fait un ingrédient d'une technique de T_P.

Ces deux éléments sont aussi valides pour les Fig. 3, 4 et 5 qui présentent les différents types de tâches identifiés dans le corpus et relatifs aux variables, aux boucles et aux instructions conditionnelles.

Pour continuer la structuration, dans la catégorie *Analyser* nous avons identifié deux genres de tâches. Ils précisent l'activité à mettre en œuvre pour mener l'analyse, et donc les techniques et les éléments du *logos* permettant de réaliser les tâches de ces catégories :

— *Identifier* : pour cette catégorie, en s'appuyant sur les éléments du logos tels que les définitions des structures ou des règles syntaxiques (délimitation des instructions exécutées dans une structure de contrôle par exemple), l'apprenant doit lire dans un code existant, l'information visée.
— *Déterminer* : pour ces types de tâches, l'apprenant va devoir agir sur le code existant (en l'exécutant ou en le simulant) pour réaliser la tâche.

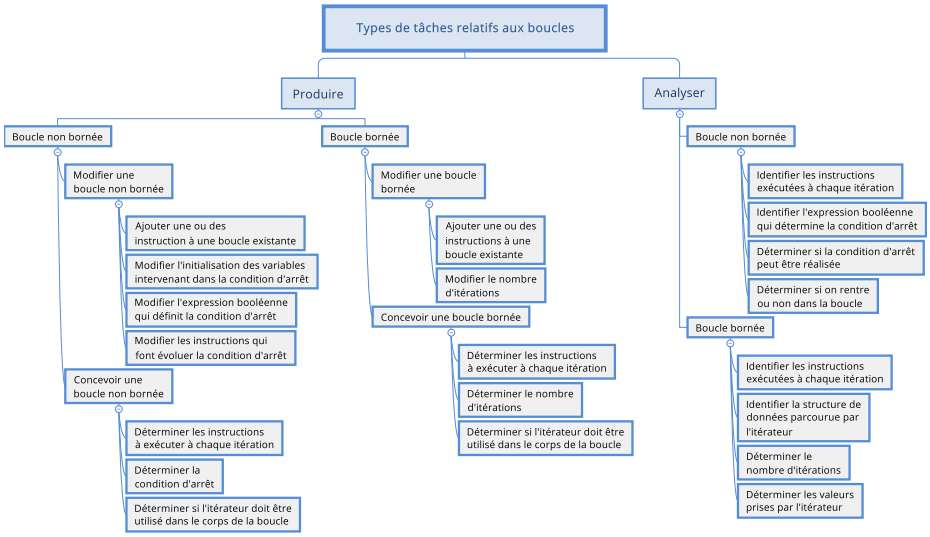


Fig. 3. Les types de tâches de la catégorie *Boucle*

Les éléments du logos mobilisés concernent d'une part l'exécution ou à la simulation d'un code et d'autre part les objets ou structures rencontrés.

Dans la catégorie *Produire* nous avons défini deux genres de tâches :

— *Modifier* : dans cette catégorie il y a déjà un code existant qu'il faut faire évoluer. Nous avons des types de tâches de ce genre dans la famille de types de tâches *Programme* et pour les deux structures de contrôle *Boucle* et *Instruction conditionnelle*. D'une manière générale la modification est guidée par l'objectif de corriger ou adapter un programme. Dans des ressources d'apprentissage elle peut être contrainte par une *limitation du nombre de lignes* ou *l'obligation ou interdiction d'utiliser une structure*.

— *Concevoir* : les types de tâches de ce genre de tâches correspondent à la situation où il n'y a encore aucun élément disponible du programme P qui doit réaliser la tâche t.

Pour le genre de tâche *Concevoir*, le type de tâches le plus générique est celui de la catégorie *Programme* qui est défini par *Concevoir un programme permettant de réaliser une tâche t*. Sa technique mobilise de nombreux autres types de tâches présents dans les arbres. Pour les structures de contrôle nous avons proposé, en dessous de chaque type de tâches global (i.e., *Concevoir une boucle bornée*), les types de tâches ingrédients des techniques qui font l'objet d'exercices spécifiques dans le corpus analysé.

Les différents types de tâches identifiés dans le corpus de ressources permettent de rendre compte de la transposition didactique réalisée dans différents moyens d'apprentissage. Même si l'étude d'autres moyens d'apprentissage permettait de mettre en évidence d'autres types de tâches, la structure globale ne

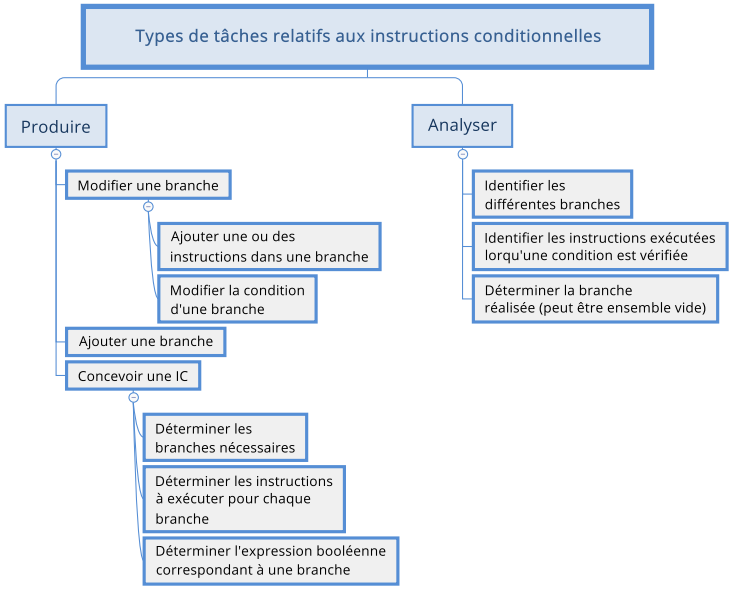


Fig. 4. Les types de tâches de la catégorie *Instruction conditionnelle*.

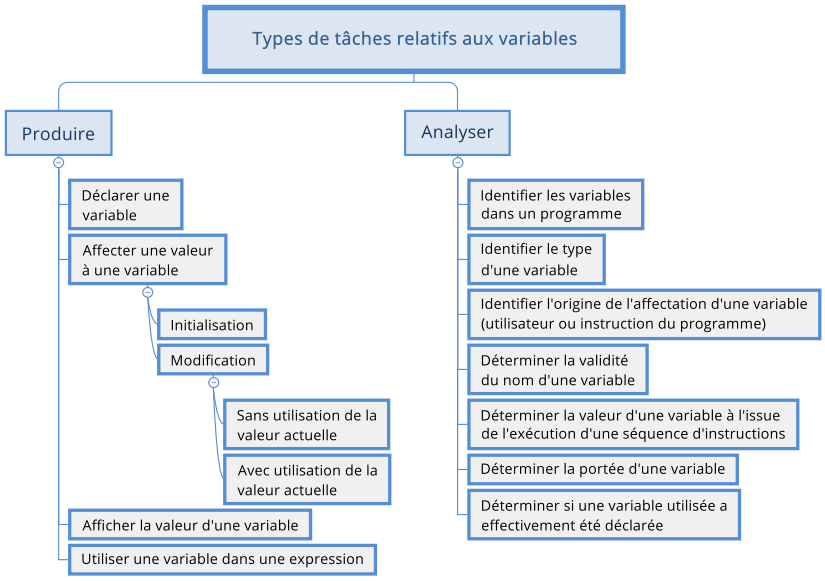


Fig. 5. Les types de tâches de la catégorie *Variable*

serait pas remise en cause et le travail présenté peut être complété. Nous présentons maintenant l'exploitation de ce travail pour analyser trois EIAH.

4 Analyse de trois EIAH

Pour répondre à (QR2), nous étudions trois EIAH qui visent l'apprentissage de Python par des débutants : AlgoPython [5], Citizen Code Python [6], et Pyrates [7] (notés AP, CCP et Py dans la suite de ce texte). Ces trois EIAH sont tous des applications Web récentes (moins de trois ans) ; les trois admettent au moins une version gratuite. AP est proposé par la société Génération 5. Il est présenté comme couvrant tous les aspects de l'algorithmique et de la programmation en Python du niveau lycée et permettant un apprentissage progressif. Il est composé de huit séries d'exercices qui traitent les notions qui nous intéressent (plus les fonctions et les listes), chaque série étant composée de 10 à 24 exercices. CCP a été développé par la société Tralalère et IOI France avec le soutien d'Amazon Future Engineer. Il propose une "initiation à la programmation pour tous" et présente la spécificité de pouvoir réaliser les tâches en Python ou en Blockly. Il est organisé en 11 saisons, chaque saison étant constituée d'un ensemble de 3 à 10 exercices. Py est un environnement développé dans le cadre d'une thèse [3] et propose huit niveaux, chacun étant une tâche de déplacement à réaliser à l'aide d'un programme en Python. Le mémo proposé permet de constater que les notions de variable, boucle et instruction conditionnelle sont abordées.

4.1 Méthode d'analyse des EIAH

Dans chaque EIAH nous n'analysons que les exercices identifiés comme portant sur les variables, les boucles ou les IC. Nous commençons par décrire les EIAH à l'aide d'une grille basée sur les types de tâches présentés dans la section précédente. Ces EIAH exploitent en effet les appels de fonctions comme outils dans les programmes à produire même si elles n'ont pas nécessairement été étudiées.

D'autre part, il arrive fréquemment qu'il existe plusieurs programmes P permettant de réaliser une tâche donnée t. Or, dans ces EIAH, il y a l'intention de faire produire un programme spécifique (par exemple qui utilise une structure de contrôle particulière). Pour tenir compte de ces intentions, nous identifions donc les exercices pour lesquels il existe plusieurs solutions (en restant dans le champ de solutions signifiantes), et laissons de côté celles bloquées par l'EIAH.

Présentation de la grille de description. Pour décrire les EIAH, nous avons construit une grille utilisée pour chaque exercice. Nous commençons par déterminer le ou les types de tâches de la catégorie *Programme* mobilisé(s). Ensuite, nous déterminons s'il existe au moins une résolution (qui a un sens) mobilisant :

5. `https://www.algopython.fr`
6. `https://www.futureengineer.fr/`
7. `https://py-rates.fr`

— uniquement des appels de fonctions prédéfinies et/ou des affectations ;
— uniquement une ou des IC (+ fonctions prédéfinies et/ou affectations) ;
— uniquement une ou des boucles bornées (+ fonctions prédéfinies et/ou affectations) ;
— uniquement une ou des boucles non bornées (+ fonctions prédéfinies et/ou affectations) ;
— une combinaison de boucles et/ou d'IC (+ fonctions prédéfinies et/ou affectations).

Pour chacune des solutions existantes, nous vérifions si cette solution est valide dans l'environnement.

Pour chaque solution valide, nous avons analysé les types de tâches mobilisés dans chacune des catégories (variable, boucles, IC). Enfin, si la solution demande l'utilisation de plusieurs structures de contrôle, nous regardons les relations existantes entre ces structures (succession ou imbrication) et combien d'instructions contiennent ces dernières. Ces éléments peuvent être considérés comme des marqueurs d'une certaine complexité des exercices.

Remplissage de la grille de description. Afin de remplir la grille nous résolvons chaque exercice. Sur la base de ces résolutions, nous identifions les différents types de tâches présents dans chaque exercice. Afin de nous assurer de la validité de notre grille nous avons procédé à un double codage à l'aveugle pour l'ensemble des exercices d'AP. Celui-ci nous a permis de constater une concordance des descriptions obtenues pour 42 exercices sur 45. Les écarts ont été réglés à l'aide du choix suivant : pour produire la description, nous décrivons uniquement les types de tâches prescrits par l'exercice. Ainsi, dans la description, nous n'associons pas à un type de tâches des ingrédients d'une de ses techniques. Par exemple si l'exercice mobilise le type de tâches *Concevoir une boucle bornée* nous ne lui associons pas le type de tâches *Déterminer les instructions à exécuter à chaque itération*, sauf dans le cas où celui-ci fait l'objet d'une question spécifique.

4.2 Description de trois EIAH

Dans cette section nous présentons quelques éléments significatifs obtenus à partir de l'analyse des grilles de description. Tout d'abord dans la Table 1 nous présentons les différents types de tâches de la catégorie *Programme* identifiés dans chaque EIAH. Les exercices présents dans les EIAH sont constitués d'une unique question, ce qui entraîne qu'un seul type de tâches de cette catégorie est présent par exercice.

On peut noter qu'aucun type de tâches n'appartient au genre de tâches *Analyser*. Dans le genre de tâches *Produire*, le type de tâches *Concevoir un programme permettant de réaliser une tâche t* est très majoritaire dans l'ensemble des EIAH, et même le seul présent dans Py. La situation est presque identique dans CCP où la seule exception est le premier exercice de plusieurs saisons qui demande de copier un code valide proposé et de l'exécuter. Dans AP, pour les premiers exercices de chaque série, il est proposé dans l'énoncé un programme très proche

Table 1. Effectif des types de tâches de la catégorie *Programme* par EIAH.

Types de tâches		AP	CCP	Py
Analyser un code existant		0	0	0
Feuille blanche	Copier et exécuter	0	6	0
	Concevoir un programme	26	43	8
Code existant : Modifier un programme P' en P qui réalise une tâche t	P' est erroné	1	0	0
	P' réalise t' proche de t	15	0	0
	P' contient des trous	3	0	0
Nombre d'exercices décrits		45	49	8

du code à produire (même nombre de lignes, même structure...), ceci représente un tiers des exercices proposés. On rencontre de manière anecdotique un exercice basé sur la correction d'un code erroné et trois exercices avec un code à trous à compléter. On peut constater que ces trois EIAH exploitent peu la diversité des types de tâches identifiés dans les moyens d'enseignement.

Dans la Table 2 nous présentons les différents types de tâches, relatifs aux variables, boucles et IC, présents dans les trois environnements.

Table 2. Effectif des types de tâches (hors catégorie *Programme*) par EIAH.

Type de tâches	AP	CCP	Py
Produire une boucle bornée	21	105	10
Utiliser la valeur d'une expression	5	0	0
Affecter une valeur à une variable, initialisation	9	24	3
Affecter une valeur à une variable, modification	0	13	0
Utiliser une variable dans une expression	6	0	3
Ajouter une ou des instructions à une boucle existante	2	0	0
Concevoir une IC	12	16	2
Déterminer les instructions à exécuter à chaque itération	1	2	0
Modifier les instructions qui font évoluer la condition d'arrêt	1	0	0
Produire une boucle non bornée	6	12	2

Quelques observations sur les différents EIAH.

- Dans CCP, il est impossible de passer un paramètre dans les fonctions de déplacement (`gauche()` et `droite()`) ce qui rend nécessaire d'utiliser une boucle dès que le déplacement dépasse deux occurrences et explique la surreprésentation du type de tâches *Produire une boucle bornée*.
- Pour contraindre l'utilisation de certaines structures de contrôle, divers moyens sont mis en œuvre. Les trois EIAH utilisent la limitation du

nombre de lignes de code. CCP et AP interdisent parfois explicitement l'utilisation d'une structure en renvoyant un message d'erreur spécifique, par exemple si l'on écrit `for` dans une situation où `while` est attendu. Enfin, Py est le seul environnement, parmi ceux étudiés, qui exploite l'aléatoire dans la tâche t à réaliser pour "forcer" l'utilisation d'une structure plutôt qu'une autre puisque le code produit doit fonctionner pour une tâche t non totalement connue à l'avance.

— La nature des exercices, déplacement d'un robot ou d'un avatar, dans CCP et Py, amène à devoir utiliser massivement des fonctions de déplacement, éventuellement avec paramètres, avant que la fonction ait été étudiée comme objet.

Éléments de synthèse. D'une manière globale on peut constater que les trois EIAH se situent plutôt dans une approche où c'est l'action qui prime et que l'articulation avec le bloc du *logos* (globalement peu présent) est laissée à la charge de l'utilisateur qui doit aller chercher l'information et se débrouiller pour l'articuler avec la tâche qu'il réalise. Ce constat amène à interroger le rôle dans l'apprentissage que peuvent, ou non, jouer ces EIAH. Ceci rejoint par ailleurs le questionnement posé dans [13]. Une analyse complémentaire, des trois EIAH, basée sur celle présentée dans cet article, est proposée dans [12].

5 Conclusion et perspectives

Dans cette contribution nous avons étudié la manifestation de la transposition didactique en étudiant les types de tâches présents dans un corpus de ressources d'apprentissage de la programmation. Nous avons proposé une structuration de ces types de tâches qui a été exploitée pour décrire et analyser trois EIAH d'apprentissage de Python.

Pour conclure, nous proposons trois perspectives. Premièrement, prolonger le travail initié dans cet article et dans [12], en réalisant un recensement et une étude systématique des EIAH se positionnant sur le même segment que ceux étudiés.

Deuxièmement, exploiter ces travaux comme guide pour la production ou l'évolution d'EIAH d'apprentissage de la programmation (diversification des tâches, proposition de parcours et de rétroactions, etc.).

Troisièmement, avec des finalités plus larges que l'unique étude d'EIAH, poursuivre la modélisation du savoir ébauchée dans cet article en complétant les praxéologies (en couverture en ajoutant différentes praxéologies, en profondeur en raffinant certains types de tâches, en complétude en ajoutant les techniques et le bloc du *logos*) et en intégrant un travail sur les erreurs. Ceci se faisant tout d'abord avec les praxéologies institutionnelles, puis en étudiant les praxéologies mises en œuvre par les apprenants. L'identification des praxéologies mises en œuvre peut notamment s'appuyer sur une étude des traces d'activités des utilisateurs, en particulier les différents codes successifs produits pour parvenir à

la réponse. L'étude et la compréhension des écarts entre praxéologies institutionnelles et praxéologies des apprenants est à la fois un enjeu didactique et de formation des enseignants.

Références

1. Modulo - Catalogue de ressources informatique. https ://apprendre.modulo-info.ch/ (2022)
2. Bays, S. : Numérique et sciences informatiques, 1ère. Prépas Sciences, ellipses edn. (2019)
3. Branthôme, M. : Apprentissage de la programmation informatique à la transition collège-lycée. Revue STICEF **28**(3) (2021). https://doi.org/10.23709/STICEF.28.3.1, `http://sticef.org/num/vol2021/28.3.1.branthome/28.3.1.branthome.htm`, medium : pdf Publisher : STICEF Version Number : 1
4. Canu, C. : Numérique et sciences informatiques, 1ère. Ellipses edn. (2019)
5. Chaachoua, H. : T4TEL, un cadre de référence didactique pour la conception des EIAH. In : Pilet, J., Vendeira, C. (eds.) Actes du séminaire de didactique des mathématiques 2018. pp. 8–25. IREM de Paris - Université Paris Diderot, Paris (2018), `https://hal.archives-ouvertes.fr/hal-02421410/document`
6. Chaachoua, H., Crisci, R., Tchounikine, P. : Un modèle praxéologique de référence pour des praxéologies mixtes dans des tâches de programmation. In : Florensa, I., Ruiz Munzon, N. (eds.) Pré-actes de CITAD7. pp. 361–372. Barcelona, Spain (2022)
7. Chevallard, Y. : La transposition didactique. Grenoble : La pensée sauvage (1985)
8. Chevallard, Y. : L'analyse des pratiques enseignantes en théorie anthropologique du didactique. Recherches en Didactique des Mathématiques **19**(2), 221–265 (1999)
9. Couderette, M. : Enseignement de l'algorithmique en classe de seconde : une introduction curriculaire problématique. Annales de Didactique et de Sciences Cognitives. Revue internationale de didactique des mathématiques (21), 267–296 (2016)
10. Divoux, C., Gouygou, C., Jospeh, N., Lassus, G., Magnier, L., Saës, G. : Cahier d'algorithmique 2de. Belin : éducation, belin/humensis edn. (2018)
11. Groz, B., Waller, E., Nancel, M., Beaudouin-Lafon, M., Marce, O. : NSI 1ère. Hachette (2021)
12. Jolivet, S., Dechaux, E., Gobard, A.C., Wang, P. : Description et analyse de trois EIAH d'apprentissage de Python. In : Actes de l'atelier APIMU : 11ème Conférence sur les Environnements Informatiques pour l'Apprentissage Humain. Brest (2023)
13. Jolivet, S., Grugeon-Allys, B. : Modélisation de parcours d'apprentissage adaptés à l'apprenant dans un EIAH. In : Florensa, I., Ruiz Munzón, N. (eds.) Pré-actes de la 7e conférence internationale sur la théorie anthropologique du didactique (CITAD7). pp. 92–106. Barcelonne (2022)
14. Sablayrolles, L., Lefevre, M., Guin, N., Broisin, J. : Design and Evaluation of a Competency-Based Recommendation Process. In : International Conference on Intelligent Tutoring Systems. pp. 148–160. Springer (2022)
15. Strock, J.M., Artaud, M. : Du logos des organisations algorithmiques dans l'enseignement secondaire. Educação Matemática Pesquisa : Revista do Programa de Estudos Pós-Graduados em Educação Matemática **21**(4) (2019), number : 4

Comment caractériser et analyser les compétences de la pensée informatique d'un jeu sérieux ?

Mathieu Muratet

Sorbonne Université, CNRS, LIP6, F-75005 Paris, France
INS HEA, 92150 Suresnes, France
mathieu.muratet@lip6.fr

Résumé. La pensée informatique est un ensemble de compétences mal maîtrisée par les enseignants de l'école obligatoire. Pour aider les enseignants à aborder ces compétences avec leurs élèves, nous développons le jeu SPY, un jeu sérieux dirigé par les compétences. Dans cet article nous conduisons une analyse du jeu SPY et nous proposons un formalisme pour décrire les compétences visées par le jeu à partir des fonctionnalités du jeu. Le formalisme proposé a été appliqué pour caractériser 21 des 26 compétences du PIAF et nous montrons comment ces caractérisations permettent de construire des indicateurs explicitant les compétences en jeu.

Mots-clés : Jeu sérieux, pensée informatique, modélisation, compétence, analyse automatique.

Abstract. Computational thinking is a discipline poorly mastered by elementary school teachers. To help teachers address these competencies with their students, we develop the SPY game, a serious game driven by competencies. In this paper we conduct an analysis of SPY and we propose a formalism to describe the competencies targeted by the game from the game mechanics. The proposed formalism has been applied to describe 21 of the 26 competencies of the PIAF conceptual framework and we show how these descriptions allow us to build indicators explaining the competencies on the line.

Keywords: Serious game, computational thinking, modeling, competency, automatic analysis.

1 Introduction et positionnement

Depuis maintenant quelques années, l'informatique est une discipline des programmes scolaires [1], où l'ensemble des compétences attendues est appelée « pensée informatique ». Selon Wing [2], la pensée informatique met en jeu un répertoire de cinq capacités cognitives : (1) la pensée algorithmique, (2) l'abstraction, (3) l'évaluation, (4) la décomposition, et (5) la généralisation. Sur la base de cette définition Parmentier *et al.*

[3] proposent un découpage fin des compétences de la pensée informatique et algorithmique pour l'enseignement fondamental (le PIAF). Ils proposent également des scénarios pédagogiques pour aider les enseignants à transmettre ces compétences à leurs élèves. Cette question de l'appropriation de ces compétences par les enseignants de l'école primaire est complexe à résoudre. En effet, enseigner cette nouvelle discipline demande un investissement particulièrement important de leur part [4] notamment en raison d'un manque de formation (initiale et continue). Ainsi, par manque de maîtrise de ces compétences les enseignants se trouvent en difficulté pour construire leurs propres scénarios pédagogiques ou juger de la pertinence d'outils pour faire travailler ces compétences à leurs élèves [5].

Nous faisons l'hypothèse que même si les enseignants de l'école primaire ne maîtrisent pas les compétences relatives à la pensée informatique, ils sont capables d'en comprendre les enjeux pour leurs élèves lorsqu'ils sont présentés dans des référentiels ; et qu'un outil basé sur ces référentiels serait lisible pour les enseignants.

Notre proposition est la suivante : Proposer un jeu sérieux d'apprentissage dirigé par les compétences qui assistera les enseignants dans le développement de sessions d'apprentissage de la pensée informatique avec leurs élèves.

Aujourd'hui, plusieurs dizaines de jeux sérieux d'apprentissage sur le thème de l'informatique existent [6, 7] mais très peu explicitent comment les compétences sous-jacentes sont travaillées et encore moins fournissent des outils à destination des enseignants pour les aider à adapter les jeux à leur contexte [8].

La problématique générale est la suivante : comment aider des enseignants non familiers à la pensée informatique à repérer dans des situations de jeu (qu'ils pourraient eux même créer) les compétences en jeu ?

Cette problématique est déclinée en deux questions de recherche :

- Comment formaliser les compétences d'un référentiel de compétence sur la pensée informatique à partir des fonctionnalités ludiques d'un jeu sérieux ?
- Comment exploiter ce formalisme pour analyser les niveaux du jeu sérieux et en extraire les compétences en jeu ?

Afin de répondre à ces questions nous présenterons le cadre théorique dans la section 2. Nous présenterons ensuite dans la section 3 le jeu SPY que nous utilisons dans le cadre de cette étude et nous exposerons ses fonctionnalités didactiques. Nous proposerons dans la section 4 un formalisme pour décrire des compétences de la pensée informatique à l'aide des fonctionnalités de jeu et nous illustrerons l'application de ce formalisme sur quelques compétences de la pensée informatique dans la section 5. Enfin, nous présenterons quelques résultats d'analyses réalisés sur le jeu SPY dans la section 6 avant de conclure.

2 Encrage théorique

Dès 1994, Balacheff [9] identifie le processus de « transposition informatique ». Il vise à identifier les liens entre l'univers interne du dispositif et l'univers externe (dans lequel se trouve l'utilisateur) au travers d'une interface. Cette interface joue un rôle particulier

lorsqu'elle devient « une référence pour l'utilisateur relativement à laquelle la connaissance est construite » [9]. Ainsi, les choix de conception d'une interface influencent le processus de construction de connaissances.

Concevoir un jeu sérieux d'apprentissage à métaphore intrinsèque consiste à placer l'apprentissage au cœur de la jouabilité [10]. Dans ce contexte les concepteurs de jeux sérieux proposent des mécanismes de *gameplay* en harmonie avec les contenus à enseigner. Il y a donc une relation forte entre les fonctionnalités de jeu et le savoir à enseigner, les fonctionnalités de jeu sont alors pour la plupart didactiques. La transposition informatique des apprentissages visés par le jeu sérieux n'est pour autant pas toujours explicite. Branthiome [11] développe l'idée de situations adidactiques dans lesquelles les élèves peuvent envisager des solutions initiales peu efficaces qui, avec les rétroactions du système, permettent de passer sur une procédure gagnante. Pour autant les fonctionnalités de jeu proposées sont didactiques pour permettre aux élèves de progresser dans leurs apprentissages même s'ils n'en ont pas conscience au cours du jeu.

Le joueur est donc en interaction avec le jeu sérieux et développe des schèmes d'utilisation qui transforment le jeu d'un statut d'artefact au statut d'instrument [12]. Les schèmes ainsi développés dépendent des propriétés de l'artefact interactif (dans notre cas des fonctionnalités de jeu disponibles dans une situation donnée). La théorie de la médiation sémiotique développe l'idée du potentiel sémiotique d'un artefact [13]. Ce potentiel est défini par un double lien « qui peut s'établir entre i) un artefact et les significations personnelles émergeant de son utilisation finalisée ; ii) cet artefact et les significations mathématiques évoquées par son usage, reconnaissables comme mathématiques par un expert » [13]. C'est ce second lien qui nous intéresse dans nos travaux où nous cherchons à caractériser les liens entre un jeu sérieux (l'artefact) et les signes issus de l'usage, reconnaissables comme relevant du savoir, dans notre cas la pensée informatique.

3 Analyse des fonctionnalités du jeu SPY

SPY[1] est jeu sérieux d'apprentissage sur le thème de la pensée informatique. Il a été conçu pour un public d'élèves de cycle 3 (CM1, CM2, 6ème). C'est un projet open source[2] développé par Sorbonne Université. Le principe du jeu est de programmer un agent à l'aide d'actions afin de l'aider à sortir d'un labyrinthe (se déplacer du point de départ au point d'arrivée du labyrinthe). Ces actions sont représentées sous forme de blocs que le joueur doit agencer en séquences exécutées par l'agent.

La mobilisation des compétences de la pensée informatique repose sur un ensemble de caractéristiques de jeu. Nous allons détailler ici l'analyse de cet artefact afin d'en identifier tous les ressorts ludiques et étudier les liens avec le savoir à enseigner : la pensée informatique.

[1] SPY : https://spy.lip6.fr, accédé le 09/01/2023

[2] Code source du jeu SPY : https://github.com/Mocahteam/SPY, accédé le 09/01/2023

3.1 Blocs de programmation

Les blocs de programmation sont les éléments fondamentaux de SPY dans la mesure où ils constituent les briques de base composant les programmes exécutés par les agents. Les blocs de programmation sont répartis en quatre catégories :

Les blocs d'action : Ces blocs permettent de définir les actions pouvant être réalisées par les agents à savoir : « Avancer », « Pivoter à gauche », « Pivoter à droite », « Attendre », « Activer un terminal » et « Faire demi-tour ». A noter que toutes ces actions sont atomiques à l'exception de l'action « Faire demi-tour » qui peut être décomposée par deux actions « Pivoter à droite » ou deux actions « Pivoter à gauche ».

Les blocs de contrôle : Ces blocs permettent de contrôler les blocs d'action à exécuter à savoir : « Si Alors », « Si Alors Sinon », « Répéter n fois », « Tant que » et « Répéter indéfiniment ».

Les capteurs : Ces blocs donnent des informations sur l'environnement avoisinant les agents. Les capteurs renvoient des valeurs booléennes qui peuvent être exploitées dans les blocs de contrôle « Si Alors », « Si Alors Sinon » et « Tant que ». Les capteurs permettent à l'agent de savoir si un mur se trouve en face de lui, à sa gauche ou à sa droite ; si un passage se trouve en face de lui, à sa gauche ou à sa droite ; si une sentinelle, une zone surveillée ou une porte se trouve en face de lui ; et si un terminal ou une sortie se trouve sur sa position.

Les opérateurs : Ces blocs permettent de combiner les capteurs. Nous retrouvons les classiques opérateurs booléens : « Non », « Ou » et « Et ». On notera que certains capteurs peuvent être exprimés par d'autres capteurs à l'aide des opérateurs, par exemple « Mur en face » est équivalent à « Non Passage en face ».

Limitation des blocs de programmation : SPY donne la possibilité de paramétrer pour chaque niveau quels blocs de programmation seront disponibles et en quelle quantité. Cette fonctionnalité peut être utile à la fois pour réduire la complexité d'un niveau ou au contraire l'augmenter. Par exemple, un niveau d'introduction pourra ne contenir que les blocs utiles à sa résolution, évitant ainsi au joueur de devoir choisir parmi des blocs potentiellement inutiles. Inversement limiter l'accès à certains blocs (ou leur quantité) peut être utile pour forcer le joueur à utiliser certains blocs (par exemple, demander de faire avancer un agent de plusieurs cases en limitant le nombre de bloc « Avancer » à 1 et en donnant accès au bloc de contrôle « Répéter n fois »).

3.2 Zone de programmation

La zone de programmation est la seconde fonctionnalité fondamentale de SPY. Elle accueille les blocs de programmation permettant de construire les solutions aux problèmes posés.

Par défaut, chaque agent est associé à une zone de programmation. Il est cependant possible de rompre cette association afin de demander au joueur de le faire. Dans ce cas, le joueur doit indiquer dans la zone de programmation le nom de l'agent auquel elle est associée afin que le programme lui soit envoyé lors du lancement de l'exécution.

Une zone de programmation peut être préconstruite. Le programme ainsi proposé au joueur peut être complet, partiel ou bogué. Le joueur devra en conséquence la compléter ou la corriger le cas échéant.

Enfin, SPY offre la possibilité de ne proposer au joueur qu'une seule zone de programmation pour plusieurs agents. Ceci permet de construire des niveaux où le joueur devra imaginer un unique programme permettant de contrôler plusieurs agents pouvant être dans des labyrinthes différents. Le joueur pourra ainsi se confronter à la réalisation de solutions plus génériques.

3.3 Glisser/déposer

Le glisser/déposer est la troisième fonctionnalité fondamentale dans le jeu SPY, car elle permet au joueur de glisser/déposer de nouveaux blocs dans la séquence d'action à exécuter par l'agent. Il est néanmoins possible de désactiver cette fonctionnalité, dans ce cas le niveau doit proposer un ou plusieurs programmes préconstruits que le joueur pourra exécuter sans pouvoir les modifier.

3.4 Obstacles

Pour résoudre les différents niveaux du jeu, le joueur doit programmer un ou plusieurs agents pour les aider à atteindre une position particulière tout en évitant des obstacles. Ces obstacles ont été conçus avec une intention didactique :

- Des sentinelles surveillent des zones du labyrinthe. Si un agent se trouve sur l'une de ces zones surveillées, l'agent est détecté ce qui met fin à la partie. Le joueur doit donc programmer l'agent pour qu'il évite ces zones surveillées. Par ailleurs, comme pour les agents, les sentinelles peuvent contenir un programme préconstruit ce qui rend les zones surveillées dynamiques. Le joueur peut donc sélectionner chaque sentinelle du jeu, observer le programme qui la compose, comprendre cette séquence d'action, anticiper les mouvements de la sentinelle et programmer son agent en conséquence.
- Des portes peuvent être ouvertes et fermées à l'aide de terminaux. Cette mécanique a été introduite pour engager un processus de résolution en étapes. Lorsqu'une porte bloque le passage, le joueur doit décomposer sa solution en sous-étapes (objectif 1 : activer le terminal pour ouvrir la porte ; et objectif 2 : atteindre la sortie). Les portes permettent également de manipuler des objets dont l'état peut changer (ouvert ou fermé).

3.5 Contrôle de l'exécution des programmes

Lorsque le joueur souhaite tester sa solution il clique sur un bouton pour lancer la simulation. Chaque zone de programmation envoie son programme à l'agent ciblé s'il existe et chaque agent/sentinelle exécute ses actions en parallèle. Le joueur peut suivre l'exécution des programmes en observant les agents/sentinelles bouger dans la scène

en fonction des actions en cours d'exécution (mises en évidence dans le panneau d'exécution des programmes). Le joueur peut mettre en pause la simulation à tout moment et exécuter les programmes pas à pas.

3.6 Nombre d'exécutions

Trouver la solution à un problème en un seul coup (trouver un programme qui permet de déplacer l'agent directement du point de départ au point d'arrivée) peut être une tâche complexe notamment sur des niveaux contenant des sentinelles en mouvement. SPY permet donc de résoudre un niveau en plusieurs coups. Le joueur peut définir une première séquence de blocs, l'exécuter, observer la nouvelle situation, ajouter de nouveaux blocs, exécuter, observer… Il est donc possible de construire le programme solution étape par étape.

Cependant, il peut être pertinent de limiter ce nombre d'exécutions pour amener progressivement le joueur à anticiper plusieurs actions en avance. Il est donc possible pour chaque niveau de SPY de définir le nombre autorisé d'exécutions pour le résoudre.

3.7 Brouillard et texte d'introduction

Dans un niveau classique de SPY le joueur a une vue omnisciente de la situation de jeu ce qui lui permet de planifier ses actions en vue d'atteindre l'objectif du niveau décrit dans les textes d'introduction. Le brouillard permet de modifier cette règle en cachant la vue d'un agent à son entourage proche. Dans ce cas, le texte d'introduction présentant l'objectif de la mission joue un rôle fondamental, car il devra contenir des indices permettant au joueur de trouver la solution. Par exemple, l'algorithme peut être donné dans le texte d'introduction en langage naturel et le joueur doit le traduire à l'aide du langage formel du jeu.

3.8 Synthèse des fonctionnalités de jeu de SPY

L'analyse des différentes fonctionnalités de SPY nous montre que chacune d'elles est didactique. Certaines sont évidentes comme les blocs de contrôle qui permettent de manipuler les notions de programmation associées, d'autres sont cachées comme l'association de programmes aux sentinelles qui demandera au joueur de comprendre une séquence d'action et d'anticiper son exécution, ou l'utilisation d'une seule zone de programmation pour contrôler deux agents qui forcera le joueur à généraliser sa solution.

D'un point de vue macro, les différentes fonctionnalités de SPY permettent de couvrir les compétences de la pensée informatique telles que définies par Wing [2]. Le joueur doit observer et modéliser la simulation (abstraction), décomposer sa stratégie en sous-étapes (décomposition), déterminer la meilleure solution (évaluation), planifier les actions à réaliser (pensée algorithmique) et réutiliser et adapter des solutions précédentes sur de nouveaux problèmes (généralisation). Mais comment identifier si telle ou telle compétence est en jeu dans un niveau donné ? Quelles influences les combinaisons de fonctionnalités ont-elles sur les compétences en jeu ?

4 Lier les compétences aux fonctionnalités de jeu

Pour décrire une compétence nous définissons des contraintes sur les fonctionnalités de jeu que nous évaluons sous la forme d'une expression booléenne. Si cette règle est vraie car l'ensemble des contraintes est satisfait, alors nous considérons que la compétence est en jeu dans le niveau analysé.

Les règles et contraintes définies pour chaque compétence s'appuient sur la structure des niveaux décrite au format XML. Le Table 1 présente les balises que nous référençons dans cet article. Un modèle de niveau complet et commenté est donné en [14].

Table 1 : Description des principales balises structurant un niveau SPY

<dragdropDisabled>	Si présente, désactive la fonctionnalité de glisser/déposer
<blockLimit blockType="X" limit="Y">	Si présente, définit la quantité Y de blocs de type X disponibles dans l'inventaire (si Y = -1 bloc en quantité illimitée).
<player inputLine="X">	Définit un agent contrôlé par le joueur. Cet agent écoute le canal de communication X.
<enemy inputLine="X">	Définit une sentinelle. Cette sentinelle écoute le canal de communication X.
<script outputLine="X" editMode="Y" type="Z">	Définit une zone de programmation qui enverra son contenu sur le canal de communication X (voir balise <player> et <enemy>). La propriété **editMode** indique si le joueur peut changer ou non le canal de communication. La propriété **type** indique si cette zone de programmation contient un programme préconstruit optimal, non optimal, bogué ou indéfini.

Une contrainte est un filtre de balises XML qui exclue de l'ensemble des balises définissant un niveau toutes les balises ne respectant pas cette contrainte. Chaque contrainte peut être précisée à l'aide d'un ensemble de paramètres. Le Table 2 décrit l'ensemble des paramètres qui peuvent être appliqués à une contrainte.

Table 2 : Description des contraintes possibles

$$TA^{U} \in ensemble\ de^{S}\ balise^{S}\ XML^{L}\ de\ SP^{Y}$$
$$ATTR^{R} \in ensemble\ de^{S}\ attribut^{S}\ valide\ pour\ TA^{G}$$
$$OP \in \{=, <, \leq, >, \geq, \neq\}$$
$$VA^{L} \in \mathbb{N}$$

TAG	Filtre les balises TAG
TAG ATTR OP VAL	Filtre les balises TAG qui ont un attribut ATTR égal/différent… à une valeur entière
TAG ATTR include SET	Filtre les balises TAG qui ont un attribut ATTR dont la valeur est incluse dans un ensemble de valeurs (SET)
TAG ATTR sameValue TAG_2 $ATTR_2$	Filtre les balises TAG qui ont un attribut ATTR dont la valeur est égale à la valeur de l'attribut $ATTR_2$ d'une balise TAG_2
TAG hasChild	Filtre les balises TAG qui contiennent au moins une balise fille

L'identification d'une compétence repose alors sur l'évaluation d'un ensemble de contraintes. Nous donnons dans la section suivante quelques exemples de contraintes que nous avons définies pour caractériser quelques compétences du PIAF.

5 Application du formalisme aux compétences du PIAF

Comme nous l'avons présenté en introduction, nous choisissons de baser notre travail sur le référentiel PIAF. Il est focalisé sur la pensée informatique et propose six compétences principales. Chacune de ces compétences est découpée en un ensemble de sous-compétences pour un total de 26 sous-compétences [15].

Les six compétences principales sont : C1 - Définir des abstractions / généraliser ; C2 – Composer / décomposer une séquence d'actions ; C3 - Contrôler une séquence d'actions ; C4 - Évaluer des objets ou des séquences d'actions ; C5 - Manipuler des représentations formelles ; et C6 - Construire une séquence d'actions de manière itérative. Dans la suite de l'article nous faisons référence aux compétences du référentiel PIAF de la manière suivante : CX.Y signifie que la sous-compétence Y de la compétence X est en jeu, par exemple C2.4 désigne la quatrième sous compétence de C2.

Nous ne développons pas ici de manière exhaustive la caractérisation de toutes les compétences du PIAF. L'ensemble des règles et contraintes définies pour caractériser les compétences du PIAF en fonction des fonctionnalités de jeu est accessible en [16]. Néanmoins, nous présentons quelques exemples d'application de notre formalisme sur une sélection de trois compétences du PIAF.

5.1 C1.1 : Nommer des objets et séquence d'actions

La compétence C1.1 est définie ainsi : « Être capable de donner des noms à des objets, des actions et des séquences d'actions » [15, p. 2].

Dans SPY nous considérons que cette compétence est en jeu lorsque le joueur associe une zone de programmation à un agent. En effet, dans ce cas il doit nommer son programme afin qu'il soit interprété par le bon agent.

Nous caractérisons cette compétence à l'aide de la contrainte « **TAG ATTR OP VAL** » qui est instanciée ainsi : « **script editMode = 2** ». Cette contrainte signifie : « filtre les balises "script" qui ont un attribut "editMode" "égal" à la valeur 2 ».

L'expression booléenne qui définit la règle complète est alors « C1.1 = Card(script editMode = 2) > 0 » qui signifie : « La compétence C1.1 est en jeu dans le niveau si le cardinal de la contrainte est supérieur à 0 », autrement dit « si le niveau contient au moins une zone de programmation nommable ».

5.2 C1.5 : Prédire le résultat d'une séquence d'actions

La compétence C1.5 est définie ainsi : « Être capable de dire, à partir d'une séquence d'actions, ce qui se passera si elle est exécutée. Contrairement à la compétence 1.4, cette compétence consiste à fournir une prédiction sans exécuter réellement la séquence d'actions » [15, p. 5].

Dans SPY, cette compétence est travaillée lorsqu'un niveau contient une sentinelle préprogrammée. Dans ce cas le joueur doit anticiper les déplacements de la sentinelle sans pouvoir exécuter son programme. Il devra proposer une première solution pour voir les mouvements de la sentinelle et vérifier ses hypothèses.

Pour décrire cette compétence, il est nécessaire de vérifier dans un niveau qu'une zone de programmation est associée à une sentinelle et qu'elle contient au moins une action.

La règle est décrite à l'aide d'un double paramètre : (P1) « **TAG ATTR** sameValue **TAG$_2$ ATTR$_2$** » d'une part et (P2) « **TAG** hasChild » d'autre part.

Le premier paramètre permet de s'assurer qu'une sentinelle (balise « enemy ») écoute le même canal de communication (attribut « inputLine ») que celui d'une zone de programmation (balise « script » et attribut « outputLine »). P1 est donc instancié ainsi : « **enemy inputLine** sameValue **script outputLine** ».

Le second paramètre permet de s'assurer que la zone de programmation (balise « script ») contient bien au moins une action. P2 est donc instancié ainsi : « **script** hasChild ».

L'expression booléenne qui définit la règle complète est alors l'intersection des deux paramètres : « C1.5 = Card(P1 ⋂ P2) > 0 ».

5.3 C2.1 : Ordonner une séquence d'actions pour atteindre un objectif

La compétence C2.1 est définie ainsi : « Étant donné une liste non ordonnée d'actions et un but, être capable de combiner ces actions dans un ordre valide pour construire une séquence qui permet d'atteindre ce but. […] Ainsi, l'apprenant n'a pas besoin d'identifier toutes les parties nécessaires pour atteindre l'objectif, mais seulement de les mettre dans le bon ordre » [15, p. 7].

Dans SPY, il s'agit d'une situation où il est proposé au joueur uniquement les blocs utiles à la résolution du problème. Le joueur doit alors simplement les combiner dans le bon ordre pour résoudre le niveau.

Pour décrire cette compétence, il est nécessaire dans un niveau de respecter les contraintes suivantes : (R1) le glisser/déposer est activé pour permettre au joueur de combiner les actions ; (R2) aucun bloc n'est disponible en quantité illimitée ; et (R3) il y a au moins une action disponible dans l'inventaire.

La règle R1 est décrite à l'aide de la contrainte « **TAG** » qui est instanciée avec la balise « **dragdropDisabled** ». Nous souhaitons vérifier que le glisser/déposer est actif et donc que la balise « dragdropDisabled » est absente dans la description du niveau. R1 est donc définie ainsi « Card(dragdropDisabled) = 0 ».

La règle R2 est décrite à l'aide de la contrainte « **TAG ATTR OP VAL** » qui est instanciée ainsi : « **blockLimit limit = -1** ». Cette contrainte signifie : « filtre les balises "blockLimit" qui ont un attribut "limit" "égal" à la valeur -1 ». Nous souhaitons vérifier qu'il n'y a pas de bloc en quantité illimitée, R2 est donc définie ainsi « Card(blockLimit limit = -1) = 0 ».

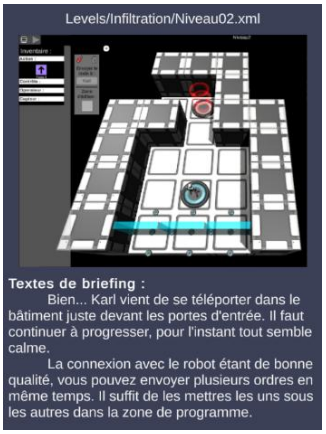
La règle R3 est décrite à l'aide d'un double paramètre : (P1) « **TAG ATTR OP VAL** » d'une part et (P2) « **TAG ATTR** include **SET** » d'autre part. Le premier permet d'identifier si un bloc est disponible dans l'inventaire (« **blockLimit limit > 0** ») et le second si ce bloc est un bloc d'action (« **blockLimit blockType** include **{Forward,TurnLeft,TurnRight,Wait,Activate,TurnBack}** »). Le cardinal de l'intersection de ces deux ensembles compte le nombre de blocs, de type action, disponibles dans l'inventaire. R3 est donc définie ainsi : « Card(P1 ⋂ P2) >=1 ».

L'expression booléenne qui définit la règle complète est alors : « C2.1 = R1 && R2 && R3 ».

6 Résultats

Nous avons illustré de manière détaillée la caractérisation de trois compétences du PIAF à l'aide du formalisme proposé. Sur l'ensemble des 26 compétences du PIAF, 21 sont applicables à SPY et ont été décrites. Outre les compétences du PIAF, nous avons confronté notre formalisme à un autre référentiel, nous avons ainsi décrit les 5 niveaux de la compétence « 3.4 Programmer » du domaine 3 « Création de contenus » du CRCN [17, p. 13]. Enfin nous avons exploité le formalisme pour exposer l'ensemble des fonctionnalités ludiques de SPY (dit référentiel SPY). L'ensemble de ces caractérisations sont accessibles en [16].

Nous pouvons ainsi analyser chaque niveau au regard des différents référentiels. Fig. 1 illustre l'analyse d'un niveau avec les 3 référentiels. Cette visualisation permet à l'enseignant d'obtenir des informations en fonction du référentiel de son choix en vue de déterminer si le niveau est un candidat potentiel pour intégrer sa scénarisation pédagogique.



Référentiel PIAF : C2.1 – Ordonner une séquence d'actions pour atteindre un objectif C2.3 – Créer une séquence d'actions pour atteindre un objectif simple C2.6 – Décomposer des objectifs en sous-objectifs plus simples C6.1 – Vérifier si une séquence d'actions atteint un objectif donné
Référentiel CRCN : Niveau 1
Référentiel SPY : F1 – Résoudre un problème en plusieurs étapes F10 – Blocs limités F11 – Bloc action « Avancer »

Fig. 1. Analyse d'un même niveau avec les différents référentiels

Nous pouvons aussi exploiter ces caractérisations pour analyser l'évolution de la complexité d'un scénario de jeu. SPY contient deux scénarios : un original à SPY intitulé « Infiltration » composé de 20 niveaux et une réplique du jeu BlocklyMaze[3] composé de 10 niveaux. Les enseignants peuvent également construire leurs propres scénarios en les composants à partir de la banque de niveaux existante. Nous avons utilisé le référentiel du PIAF pour caractériser la dimension didactique des différents scénarios

[3] BlocklyMaze : https://blockly.games/maze, accédé le 13/01/2023

et le référentiel SPY pour caractériser leur dimension ludique. Le cumul des compétences et fonctionnalités ludiques dans chaque niveau donne un indicateur sur sa complexité (voir Fig. 2). Nous observons ainsi l'évolution progressive de la complexité des deux scénarios et l'apparition des différentes compétences au cours des deux scénarios. Cette visualisation renvoie aux travaux de Carron *et al.* [18] qui analysent manuellement chaque niveau de jeu sur les dimensions ludiques et pédagogiques. Dans notre cas, l'analyse des niveaux est automatisée.

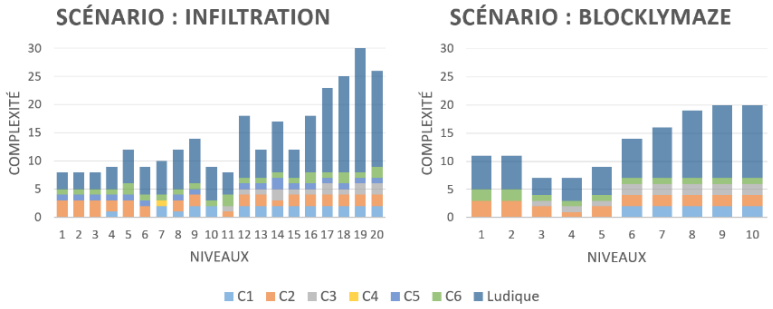


Fig. 2. Analyse des deux scenarios du jeu SPY

7 Conclusion et perspectives

Dans cet article nous avons réalisé une analyse du jeu SPY en explorant les fonctionnalités principales du jeu par rapport à des référentiels de compétences sur la pensée informatique. Nous avons proposé un formalisme pour décrire des compétences par combinaison de fonctionnalité de jeu. Nous avons ensuite exploité ce formalisme pour proposer une caractérisation des 21 compétences PIAF en jeu dans SPY, des 5 niveaux de la compétence 3.4 du CRCN et des 35 fonctionnalités ludiques de SPY. L'ensemble des résultats sont présentés en [16]. Nous avons ainsi montré que le formalisme proposé est indépendant d'un référentiel donné. Nous avons illustré dans cet article un extrait de ces résultats à l'aide des compétences C1.1, C1.5 et C2.1 du PIAF.

Enfin, nous avons montré comment ces caractérisations peuvent être exploitées pour construire des indicateurs sur des niveaux existants et fournir aux enseignants des informations *micros* sur les compétences en jeu dans un niveau ou des informations *macros* sur l'évolution de la complexité d'un scénario de jeu.

Actuellement le jeu SPY ne contient pas d'éditeur de niveaux. La création ou la modification de niveaux reste peu accessible (éditer à la main les fichiers XML). Une perspective pour ce travail est de proposer un tel éditeur, afin de permettre aux enseignants de créer leurs propres niveaux de jeu et d'en obtenir une analyse automatique à partir du référentiel de leur choix.

Une autre piste de recherche que nous avons identifiée consisterait à exploiter la topologie des labyrinthes comme une information susceptible d'être didactique et donc d'être accessible dans le formalisme. Ainsi, en inversant la chaîne, nous pourrions envisager d'automatiquement créer de nouveaux niveaux à partir d'une sélection de compétences visées.

References

1. Baron, G-L., Drot-Delange, B.: L'informatique comme objet d'enseignement à l'école primaire française ? Mise en perspective historique. Revue française de pédagogie Recherches en éducation, 51–62 (2016).
2. Wing, J. M.: Computational thinking. Communications of the ACM, 49(3), 33–35 (2006).
3. Parmentier, Y., Reuter, R., Higuet, S., Kataja, L., Kreis, Y., Duflot-Kremer, M., Laduron, C., Meyers, C., Busana, G., Weinberger, A., Denis, B.: PIAF: developing computational and algorithmic thinking in fundamental education. In: EdMedia+ Innovate Learning, Association for the Advancement of Computing in Education (AACE), pp. 315–322 (2020).
4. Kradolfer, S., Dubois, S., Riedo, F., Mondada, F., Fassa, F.: A sociological contribution to understanding the use of robots in schools: the thymio robot. In: International Conference on Social Robotics, Springer, Cham, pp. 217–228 (2014).
5. Poirson, B.: Appuis et obstacles à l'apprentissage de la pensée informatique à l'école primaire. Education (2020).
6. Lindberg, R. S., Laine, T. H., Haaranen, L.: Gamifying programming education in K-12: A review of programming curricula in seven countries and programming games. British Journal of Educational Technology 50(4), 1979–1995 (2019).
7. Miljanovic, M., Bradbury, J.: A Review of Serious Games for Programming. In: Proc. of the 4th Joint Conference on Serious Games (JCSG 2018), Darmstadt, Germany (2018).
8. Saddoug, H., Rahimian, A., Marne, B., Muratet, M., Sehaba, K., Jolivet, S.: Review of the Adaptability of a Set of Learning Games Meant for Teaching Computational Thinking or Programming in France. In: Proc. of the 14th International Conference on Computer Supported Education, vol. 1: GonCPL, pp 562–569 (2022).
9. Balacheff, N.: La transposition informatique, un nouveau problème pour la didactique. In: Artigue M. et al. (eds) Vingt ans de didactique des mathématiques en France, pp. 364–370 (1994).
10. Marne, B., Huynh-Kim-Bang, B., Labat, J.-L.: Articuler motivation et apprentissage grâce aux facettes du jeu sérieux. In: EIAH 2011, Mons, Belgique, pp. 69–80 (2011).
11. Branthôme, M.: Conception et évaluation du jeu sérieux Pyrates : aspects didactiques dans la construction des niveaux. In: RJC EIAH, Lille, France (2022).
12. Rabardel, P.: Les hommes et les technologies, approche cognitive des instruments contemporains. Paris : Armand Colin (1995).
13. Mariotti, M. A., Maracci, M.: Les artefacts comme outils de médiation sémiotique : quel cadre pour les ressources de l'enseignant ? Ghislaine Gueudet et Luc Trouche. Ressources vives. Le travail documentaire des professeurs en mathématiques., Presses Universitaires de Rennes et INRP, pp. 91–107, Paideia (2010).
14. Modèle de niveau complet et commenté au format XML, https://github.com/Mocahteam/SPY/blob/master/Doc/LevelModel.xml, accédé le 10/01/2023
15. La liste complète des 26 compétences du référentiel PIAF et leur description : https://piaf.loria.fr/wp-content/uploads/2021/09/Examples-PIAF-FR-aout-2021.pdf, accédé le 07/01/2023
16. Description de compétences à l'aide des fonctionnalités du jeu SPY, https://github.com/Mocahteam/SPY/blob/master/Assets/StreamingAssets/Competencies/competenciesReferential.json, accédé le 10/01/2023
17. CRCN, https://eduscol.education.fr/document/20389/download, accédé le 12/01/2023.
18. Carron, T., Muratet, M., Marne, B., Yessad, A.: Analyser et représenter la progression de la difficulté d'un jeu sérieux du point de vue ludique et pédagogique. In: EIAH 2017, Strasbourg, France (2017).

EIAH2023 : 11ème Conférence sur les Environnements Informatiques pour l'Apprentissage Humain
La conférence pluridisciplinaire francophone sur la conception et l'analyse des environnements numériques pour l'éducation et la formation
12-16 juin 2023 Brest (France)

Session de communications 4.1
Recherche en EIAH : enjeux et méthodologies

Environnements informatiques pour l'apprentissage du vocabulaire : une revue systématique de la littérature

Enzo Simonnet[1,2][0009−0000−9740−5212], Mathieu Loiseau[2][0000−0002−9908−0770] et Elise Lavoué[1,2][0000−0002−2659−6231]

[1] Université Jean Moulin Lyon 3, iaelyon school of Management
[2] Univ Lyon, INSA Lyon, CNRS, UCBL, LIRIS, UMR5205, F-69621

Résumé. L'acquisition du vocabulaire est une composante essentielle de l'apprentissage d'une langue étrangère, d'où l'intérêt des chercheurs et praticiens pour les environnements informatiques pour l'apprentissage du vocabulaire. Cette revue systématique de la littérature analyse un total de 40 études menées entre 2015 et 2022 sur ces environnements. Cet article décrit les outils développés et les différentes méthodologies expérimentales utilisées, en précisant les outils de mesure. Nous mettons en lumière les effets des outils d'apprentissage du vocabulaire sur l'apprentissage, la motivation, le flow et l'engagement des apprenants. Les résultats ont quasi-unanimement démontré un effet positif des outils étudiés sur les compétences lexicales, la motivation, le flow et l'engagement des apprenants. Nous apportons des éléments de discussion sur plusieurs biais possibles, ainsi que la difficile généralisabilité des résultats obtenus, en partie imputable au choix des tests utilisés dans les expérimentations.

Mots-clés : EIAH , Apprentissage du vocabulaire , Lexique , Didactique des langues , Motivation , Engagement , Revue de littérature

Abstract Vocabulary acquisition is an essential component of foreign language learning. Computational environments for vocabulary learning are thus of interest to both researchers and practitioners. This systematic literature review analyses a total of 40 studies conducted between 2015 and 2022 on such environments. This paper provides a brief overview of the tools developed and the experimental methodologies used, including the measuring tools. We highlight the clear predominance of written vocabulary activities as well as the multimodality of the content, generally combining written, audio and pictorial content. The results almost unanimously demonstrate a positive effect of the tools studied on learners' lexical skills, motivation, flow and engagement. We discuss possible biases, as well as the difficult generalisability of the results obtained, partly due to the choice of tests used in the experiments.

Mots-clés : Vocabulary learning , Motivation , Engagement , Flow , Litterature review

1 Introduction

L'apprentissage du vocabulaire est une dimension essentielle de l'apprentissage d'une langue étrangère, au même titre que la grammaire, la phonologie ou la culture [32]. L'étude du vocabulaire est une tâche complexe, combinant de multiples compétences [32, 1], mais également fastidieuse et répétitive, susceptible de démotiver les apprenants. Afin d'aider les apprenants dans cette tâche, de nombreux environnements informatiques pour l'apprentissage du vocabulaire (EIAV [3]) ont été créés et étudiés par les laboratoires de recherche et les entreprises [40]. Plusieurs revues de la littérature et méta-analyses décrivant les EIAV ont déjà été produites. Zou *et al.* [48] se sont intéressés à l'apprentissage du vocabulaire par le jeu numérique et concluent que les jeux étudiés ont globalement amélioré les compétences lexicales, la motivation et l'engagement des participants. Quatre autres études se sont spécifiquement concentrées sur l'apprentissage du vocabulaire assisté par mobile en anglais langue étrangère [3] ,[25] ,[19] ,[40]. Enfin, deux autres études concernent les EIAV sur mobile et PC, l'une pour l'apprentissage de l'anglais langue étrangère exclusivement [13] et l'autre sans restriction de langue [45]. Ces 6 études mettent en avant un impact positif des outils sur l'apprentissage du vocabulaire, mais ne prennent pas en compte les études postérieures à 2018. C'est pourquoi nous proposons d'examiner les études plus récentes afin d'actualiser les conclusions et discuter les avancées du domaine.

Dans ce contexte, notre article a pour principal objectif d'actualiser les conclusions tirées par les études précédentes, en apportant une revue systématique des articles de recherche traitant de l'apprentissage du vocabulaire en langue étrangère. Nous abordons plus particulièrement les questions de recherche suivantes : (1) Quelles variables ont été mesurées et par quelles méthodes ? Nous apportons une vue d'ensemble des études empiriques menées et les outils expérimentés, afin ensuite d'identifier les variables susceptibles d'influencer l'efficacité des EIAV. (2) Quels sont les effets de l'utilisation des EIAV ? Nous explorons l'efficacité globale de différents types d'interventions en matière d'apprentissage du vocabulaire, de motivation, d'engagement et de flow des apprenants, que ce soit à l'intérieur ou à l'extérieur de la classe.

2 Méthode

Notre processus d'analyse systématique de la littérature est fondé sur les recommandations et méthodes décrites par Webster & Watson [41]. Nous avons défini le champ de notre étude lié aux EIAV en spécifiant une requête visant à retrouver toute la littérature correspondante. Ensuite, nous avons effectué la requête et filtré les articles hors de notre champ d'étude. La requête est la suivante : *(vocabulary OR lexicon) AND (learning OR teaching OR training) AND (application OR Web OR environment).*

3. Nous introduisons cet acronyme pour des raisons de lisibilité dans la suite du document, bien que son utilisation dans la littérature ne soit pas courante

La première partie de la requête est utilisée pour filtrer les articles traitant du vocabulaire et des lexiques. La deuxième partie permet de s'assurer que les articles sont liés à l'enseignement, « *teaching* », et l'apprentissage, « *learning* » et « *training* ». Pour la dernière partie, la combinaison des 3 mots « *application* », « *web* » et « *environnement* » permet de filtrer les articles relevant du domaine des environnements numériques. Nous avons préféré ne pas inclure les termes « *computer* » et « *mobile* » car ils généraient un nombre trop important de résultats inadéquats, en lien avec des thématiques diverses telles que le réseautage social (« *social networking* ») ou l'exploration de données.

Nous avons appliqué notre requête sur différentes bibliothèques numériques scientifiques (ACM, IEEE, ScienceDirect) et revues spécialisées du domaine (ReCALL, CALL, ILLT, CALICO et LLT). Les requêtes ont été appliquées sur les titres et les mots-clés. Nous avons décidé de restreindre la date de publication de 2015 à 2022 afin de donner la priorité aux outils et recherches les plus récentes sur l'apprentissage du vocabulaire. La généralité des termes employés dans la requête a entraîné un très grand nombre d'occurrences initiales (578) constituées d'une part conséquente de bruit.

Table 1. Nombre de papiers avant et après filtrage. Le nombre d'articles conservés est donné pour chaque étape de filtrage

Nombre d'articles restants	ACM	IEEE	ScienceDirect	ReCALL	CALL	ILLT	CALICO	LLT	Total
Requête faite le 30/06/2022	15	159	257	34	48	7	28	30	
Filtre hors sujet	15	96	84	34	48	7	28	30	
Filtre thématique	7	29	26	13	21	4	13	8	
Filtre format	2	2	9	7	10	3	5	2	40

La première étape de filtrage a consisté à exclure les articles identifiés par les bibliothèques numériques comme relevant d'un sujet différent de nos thématiques de recherche tels que le *machine learning*. En deuxième étape, nous avons examiné les titres et résumés pour exclure ceux ne correspondant pas à notre domaine d'étude. Enfin, la troisième étape a consisté à exclure les publications sous forme de posters, les ateliers, les articles courts de moins de 6 pages, les méta-analyses, les revues, les doublons, les articles non publiés et ceux non écrits en anglais. Il reste ainsi 40 articles à étudier. Parmi ceux-ci, quatre ne présentent aucune expérimentation et trois articles décrivent chacun deux expérimentations distinctes. Nous analysons donc un total de 39 expérimentations.

3 Expérimentations

3.1 Durée et Participants

Les expérimentations ont été menées sur des durées différentes homogénéisées par nos soins pour permettre la description de la distribution. La plus courte durée est d'une journée, la plus longue de 19 semaines, avec une moyenne de 33,6

jours et un écart-type de 36,4 jours. Nous les avons classées en trois groupes : courtes, moyennes et longues durées, correspondant respectivement aux durées de 1 à 7 jours (1 semaine), 8 à 31 jours (1 mois), 32 jours et plus. Les groupes sont répartis de manière assez égale puisque nous comptons 12 expérimentations courtes, 14 moyennes et 13 longues. Au cours des expérimentations, le temps d'exposition au système, c'est-à-dire le temps pendant lequel les apprenants ont utilisé le système, est très variable : la durée d'exposition la plus courte est de 12 minutes et la plus longue de 140 heures, avec une valeur médiane de 6,5 heures. Dans 26% des cas, cette durée n'est pas précisée. Parfois, il peut s'agir d'une simple omission mais la plupart du temps, la durée précise est inconnue des chercheurs lorsqu'ils laissent les sujets utiliser librement les outils en dehors de la classe [8] ,[9] ,[10] ,[21] ,[23] ,[37].

Nous avons classé les participants selon leur âge et niveau scolaire (voir tableau 2), et le pays dans lequel est mené l'étude (voir tableau 3.1). Nous constatons que le public le plus représenté est celui des étudiants (niveau universitaire), donc essentiellement jeunes adultes, que nous retrouvons dans 18 expérimentations (46%). Nous constatons également une prédominance des expériences menées en Asie (72%), où les langues les plus couramment parlées sont très différentes de la langue apprise, principalement l'anglais (voir la section « Outils » 3.2). Seul 26% (10) des études ont exprimé le niveau des apprenants en fonction du Cadre européen commun de référence pour les langues (CECRL). Toutefois, les expérimentations ont toutes en commun d'impliquer des participants non spécialistes de la langue que l'on désigne en France sous l'acronyme Lansad (Langues pour les Spécialistes des Autres Disciplines).

Table 2. Nombre d'expérimentations par niveau scolaire des participants

École primaire (4-10 ans)	Collège (11-14 ans)	Lycée (15-18 ans)	Université (18+ ans)	Adultes (hors université) (18+ ans)
8	6	3	18	4

Table 3. Nombre d'expérimentations par nationalité des participants

Taiwan	n.s.	Japon	Chine	Iran	Turquie	Malaisie	Grèce	Espagne	Indonésie	Hong Kong	Russie
11	6	5	4	4	2	2	1	1	1	1	1

3.2 Outils

Dans l'ensemble des articles, 38 outils différents ont été étudiés. L'anglais comme langue étrangère est prééminent puisque 79% (30) des outils sont dédiés à son étude exclusive ; les 8 restant permettant d'apprendre des langues diverses. 18 outils sont utilisables sur ordinateur, 22 sont des applications mobiles et 6 d'entre eux sont compatibles PC et mobile. Nous retrouvons également 4 outils [35] ,[15] ,[4] ,[22] de réalité augmentée, 2 outils de réalité virtuelle [2] ,[38], un outil utilisant des objets tangibles [26] et un outil avec capteurs de mouvements [20].

La totalité des outils présentent le vocabulaire sous sa forme écrite. La prononciation des mots, ou d'autres contenus audio tels que des discours, des histoires, des musiques (avec les paroles), sont disponibles dans 50% (19) des outils, alors que la vidéo est un support utilisé par seulement 18% (7) d'entre eux. Dans 50% (19) des cas, les images sont utilisées comme un contenu complémentaire pour l'apprentissage du vocabulaire. Une grande partie des outils utilisent des contenus multimodaux : 71% (27) combinent au moins 2 formes de médias différents et 34% (13) en combinent 3 ou 4. Parmi les nombreuses activités d'apprentissage, 48% (18) des outils proposent au moins une activité de production écrite à travers des quiz, des tests ou des exercices variés ; seuls 4 outils (11%) proposent des exercices de prononciation, bien plus complexes à mettre en place du point de vue technique ; 79% (30) des outils contiennent une activité de réception du vocabulaire (cartes mémoires, listes de mots, informations lexicales). Nous avons regroupé sous l'étiquette « ludique » tous les EIAV décrits par les auteurs des articles comme des jeux ou employant des éléments de jeu tels que des récompenses virtuelles, une narration, un jeu de rôle, ou encore un tableau des scores. Cela représente 45% (17) des outils. Enfin, seuls 4 outils [5] ,[37] ,[8] ,[21] ,[34] ,[46] ,[12] proposent des fonctionnalités permettant aux apprenants de travailler collaborativement, au moyen de fonctions de discussion, de partage de liste ou d'activités collaboratives.

4 Mesures et résultats

Dans l'ensemble des expérimentations, trois catégories de variables observées ont été identifiées. Tout d'abord, 92% (36) des expérimentations mesurent au moins un aspect de l'apprentissage du vocabulaire. Par ailleurs, la motivation est mesurée dans 8 (21%) des expérimentations : [9] ,[10] ,[14] ,[21] ,[27] ,[34] ,[43] ,[38]. Enfin, nous avons regroupé le flow et l'engagement dans une seule catégorie comptant 6 expérimentations [15] ,[23] ,[27] ,[35] ,[37] (15%). Une grande variété d'autres variables ont été analysées, telles que l'opinion des participants sur les outils, la charge cognitive induite, le comportement d'apprentissage ou l'anxiété. Ces variables n'étant étudiées que dans quelques expérimentations chacune, nous n'avons pas pu généraliser les conclusions spécifiques aux résultats obtenus.

4.1 Effet sur l'apprentissage du vocabulaire

35 des 36 expérimentations testant les compétences en vocabulaire évaluent l'impact de l'usage des environnements d'apprentissage sur au moins une compétence écrite. La dernière se concentre uniquement sur une compétence orale, la prononciation [5]. Nous trouvons également 4 expérimentations testant au moins une compétence orale (discussion, prononciation ou compréhension de conversation) en plus de compétences écrites [6] ,[5] ,[10] ,[27]. Afin de mesurer les différentes compétences lexicales, 3 méthodes ont été utilisées : (a) la création d'un test *ad hoc* spécifique (67% (24) des expérimentations) ; (b) l'utilisation de

tests standardisés (30% (10) des expérimentations), le plus utilisé étant le *Vocabulary Knowledge Scale* (VKS) [7] ; (c) l'analyse des logs d'usage des outils [37] ,[5] , [11] ,[49] ,[16]. Parmi les 36 études mesurant l'apprentissage du vocabulaire, toutes sauf une [37] ont rapporté un impact positif significatif de l'outil sur les compétences en vocabulaire des sujets. Ces résultats positifs ont toutefois des significations différentes, du fait de protocoles de recherche différents. L'évolution des variables mesurées dans 12 expérimentations (33%) dans un design quasi expérimental pré-test / post-test sans groupe de contrôle ont toutes montré une augmentation de la connaissance du vocabulaire.

Les 9 expérimentations [1] ,[6] ,[12] ,[17] ,[24] ,[31] ,[34] ,[36] ,[42] (25%) où le groupe test utilise un logiciel et le groupe témoin exécute la même activité, ou du moins une activité très similaire, sur un support non technologique, isolent l'effet du dispositif d'interaction. Elles nous renseignent sur l'impact des outils numériques sur l'apprentissage du vocabulaire par rapport aux outils non numériques, traditionnels ou sur papier. 6 expérimentations [6] ,[12] ,[17] ,[31] ,[34] ,[36] parmi les 9 ont obtenu des performances supérieures en matière d'apprentissage du vocabulaire dans les groupes tests par rapport aux groupes témoins. Deux expérimentations [42] ,[24] ont fait état d'un avantage léger mais significatif du logiciel par rapport à l'activité au format papier, l'une d'entre elles rapportant que les scores des utilisateurs de l'application étaient 7% meilleurs que ceux des participants du groupe témoin [42]. L'étude menée par Abdul Wahdi & Dzulkifli [1] a par ailleurs mis en avant une augmentation équivalente des compétences en vocabulaire après l'intervention chez les participants du groupe test bénéficiant d'un EIAV et du groupe témoin bénéficiant d'un enseignement conventionnel. Une autre catégorie d'expérimentations assez similaires comprenant 4 études [2] ,[30] , [11] ,[47] (11%) a comparé un groupe test qui a utilisé un outil numérique et un groupe contrôle qui s'est appuyé sur un apprentissage conventionnel, ce qui signifie que le groupe a assisté au cours dans des conditions écologiques. De la même manière que pour la catégorie précédente, les outils numériques montrent de meilleurs résultats sur l'apprentissage du vocabulaire. Les différents types d'outils testés ont tous permis d'accroître les compétences en vocabulaire par rapport à l'approche traditionnelle : le dispositif de réalité virtuelle [2], l'utilisation d'ordinateurs [30] ,[11] ou de mobiles [47] ont tous présenté une amélioration significativement plus forte des compétences en vocabulaire par rapport aux groupes tests bénéficiant d'un apprentissage traditionnel. Les autres expérimentations [43] ,[38] ,[22] ,[28] ,[10] ,[9] ,[49] ,[16] ,[15] ,[18] comparent l'impact de caractéristiques spécifiques telles que l'utilisation de la *VR*, la stratégie de carte mentale, les mécanismes d'autorégulation sur l'apprentissage du vocabulaire.

En résumé, en accord avec les revues et méta-analyses précédentes [3] ,[25] ,[19] ,[40] ,[13] ,[45] ,[48], les expérimentations présentées ici confirment l'hypothèse que les EIAV montrent une efficacité pour l'apprentissage du vocabulaire.

4.2 Effets sur la motivation

Différents tests ont été utilisés pour mesurer la motivation, en fonction des cadres théoriques adoptés, qui peuvent varier d'une études à l'autre : (a) des

tests *ad hoc* [38], (b) des approches qualitatives (entretiens semi-directifs [6], observations en classe [21], discussion en focus groupes [35]) ; (c) des tests standardisés existants tels que le English-Learning Motivation Questionnaire [44] ou le questionnaire sur les stratégies d'apprentissage motivées (MSLQ) [33]. La motivation est mesurée dans 18% (7) des expérimentations [21] ,[14] ,[27] ,[38] ,[9] ,[43] ,[34] parmi lesquelles 3 n'ont pas de groupe de contrôle [21] ,[14] ,[27]. Kose *et al.* [21] ont utilisé une question qualitative demandant directement aux participants à quel point ils se sentaient motivés ; les réponses indiquent que les participants se sentaient motivés lors de l'utilisation des outils. En utilisant une échelle de motivation validée [29], Have *et al.* [14] ont mis en évidence une légère augmentation de la motivation des participants à apprendre du vocabulaire. Enfin, Liu *et al.* [27] ont suivi la motivation de 100 étudiants la $5^{\text{ème}}$, la $12^{\text{ème}}$ et la $19^{\text{ème}}$ semaine de l'expérimentation et ont mesuré que celle-ci augmentait au cours des semaines. Les quatre autres expérimentations [43] ,[9] ,[34] ,[38] comparent différentes conditions. La comparaison du même outil avec et sans mécanismes d'autorégulation montre une plus grande motivation avec lesdits mécanismes [9]. Pareillement, le même logiciel affiché sur un écran d'ordinateur et sur un dispositif de réalité virtuelle a révélé des niveaux de motivation plus élevés pour le groupe ayant utilisé la *VR* [38]. Par une question qualitative, Sato *et al.* [34] ont déterminé que les participants ayant utilisé Quizlet se sentaient plus motivés que le groupe révisant des listes au format papier. Enfin, aucune différence de motivation n'a été constatée dans l'expérimentation comparant un groupe utilisant une approche basée sur la complexité cognitive et un jeu de vocabulaire situationnel classique [43].

4.3 Effets sur le flow et l'engagement

Pour mesurer l'engagement des participants, Stockwell & Liu [37] ont utilisé les traces d'interactions, entretiens, questionnaires et notes d'observation. Des questionnaires *ad hoc* ont été privilégiés dans les 4 autres expérimentations [35] ,[23] ,[27] ,[15]. L'état de flow des participants, mesuré dans 3 expérimentations [27] ,[23] ,[15], n'a été comparé ni dans des contextes expérimentaux « numériques *vs.* non numériques », ni dans des contextes expérimentaux « apprentissage conventionnel *vs.* numérique ». Une expérimentation comparant une version autogérée et une version basée sur l'exécution de tâches du même logiciel [15] a montré que l'état de flow du groupe autogéré était significativement plus élevé que celui de l'autre groupe. Li *et al.* [23] montrent que la concentration et la motivation intrinsèque affectent positivement l'apprentissage perçu et la satisfaction. De plus, la satisfaction est également affectée positivement par l'apprentissage perçu. Liu *et al.* [27] mesurent la motivation des participants et les patterns de flow dans les activités de narration numérique, mais seulement en tant que mesures intermédiaires, qui une fois combinées sont utiles pour analyser l'engagement des étudiants. Le modèle d'engagement des élèves présenté dans l'article [27] est un modèle dynamique, comprenant deux cycles de désengagement et de réengagement.

5 Discussion et conclusion

Des études antérieures sur l'apprentissage du vocabulaire assisté par des jeux sérieux ont démontré l'impact positif de ces derniers sur l'apprentissage du vocabulaire [48]. Ceci est fortement soutenu par les résultats des 18 expérimentations menées sur les jeux et outils ludifiés étudiés dans cette revue. Par exemple, 5 expérimentations [2] ,[1] ,[34] ,[36] ,[47] comparant les effets d'un jeu ou d'un outil ludique sur l'apprentissage du vocabulaire des apprenants avec les effets d'un apprentissage classique en classe rapportent unanimement un impact plus important des jeux ou des outils ludiques. De plus, une étude [10] a comparé les effets d'un outil dans une version gamifiée et non-gamifiée afin d'isoler l'impact des éléments de jeu et montre que les apprenants utilisant la version gamifiée ont nettement dépassé le groupe contrôle, en termes d'acquisition et de rétention de vocabulaire. Un autre constat concerne la taille de l'effet mesuré sur l'apprentissage du vocabulaire. Celui-ci étant plutôt faible, cela laisse supposer qu'un effet systématique et minime pourrait être causé par des facteurs externes pouvant expliquer partiellement ou totalement cet effet. Tout d'abord, le fait qu'au cours de l'expérimentation l'attention des participants soit attirée explicitement sur l'apprentissage du vocabulaire est très susceptible de déclencher une amélioration de leurs performances [18, 106]. En effet, le passage d'un apprentissage du vocabulaire implicite (c'est-à-dire se produisant occasionnellement, au cours d'autres apprentissages), à un apprentissage explicite du vocabulaire (c'est-à-dire où le vocabulaire est l'objet central de l'apprentissage) pourrait être pour une part significative dans les résultats observés. Nous pouvons également supposer que l'effet de nouveauté [39], phénomène qui éveille la curiosité d'un individu qui est confronté pour la première fois à une nouvelle expérimentation, peut entraîner une augmentation de la durée d'attention. Considérant que 69% (27) des expérimentations ont duré moins d'un mois, cet effet de nouveauté pourrait ainsi influer pour une part non négligeable sur les résultats positifs observés. Pour appuyer cette affirmation, nous notons que l'expérimentation qui n'a trouvé aucun impact de l'outil numérique [37] et celle qui a trouvé un très faible impact (une amélioration de 7% de l'apprentissage du vocabulaire lors de l'utilisation de l'outil numérique) [42] sont toutes deux des expérimentations de longue durée.

En conclusion, cette revue systématique de la littérature a illustré l'effet positif global des outils numériques pour l'apprentissage du vocabulaire sur l'apprentissage, la motivation, le flow et l'engagement. Cela vient confirmer et renforcer les conclusions des revues et méta-analyses précédentes [45] ,[13] ,[45]. Cependant, nous avons également identifié plusieurs biais potentiels et causes systématiques externes qui pourraient avoir un impact significatif sur ces résultats, notamment l'effet de nouveauté et la concentration de l'attention des participants sur le vocabulaire. C'est pourquoi nous suggérons de concevoir des protocoles d'expérimentation plus comparables, par l'emploi de tests standardisés pour l'ensemble du domaine des EIAV, afin de pouvoir généraliser les résultats obtenus et tirer des conclusions de ces expérimentations utilisant des outils variés. D'autre part, nous postulons que l'analyse des traces d'utilisation pourrait

être d'une grande utilité afin d'observer l'évolution dans le temps des variables et expliquer avec plus de finesse et une meilleur qualité interprétative les résultats obtenus par les pré-tests et post-tests. Enfin, lors du design des EIAV, nous suggérons d'accorder une place plus importante aux enseignants et aux contextes d'apprentissage universitaires. En effet, les études précédentes [3] ont insisté sur le fait que les technologies ont le plus grand impact sur l'apprentissage du vocabulaire des étudiants lorsqu'elles sont introduites et expliquées par les enseignants et que l'utilisation est accompagnée par ces derniers.

Références

1. Abdul Wahdi, E.V.F., Dzulkifli, M.A. : The Effectiveness of Computer-Assisted Instruction in Vocabulary Learning. In : ICT4M. pp. 219–226 (2018)
2. Alfadil, M. : Effectiveness of virtual reality game in foreign language vocabulary acquisition. Computers & Education **153**, 103893 (2020)
3. Alhuwaydi, A. : A Review on Vocabulary Learning-Designed MALL Applications in the EFL Context. Theory and Practice in Language Studies **12**(10), 2191–2200 (2022)
4. Alobaydi, E.K., Alkhayat, R.Y., Arshad, M.R.M., Ahmed, E.R. : Context-aware ubiquitous Arabic vocabularies learning system (U-Arabic) : A framework design and implementation. In : 2017 7th IEEE International Conference on Control System, Computing and Engineering (ICCSCE). pp. 23–28 (2017)
5. Barcomb, M., Cardoso, W. : Rock or Lock ? Gamifying an online course management system for pronunciation instruction : Focus on English /r/ and /l/. CALICO Journal **37**(2), 127–147 (2020)
6. Bashori, M., van Hout, R., Strik, H., Cucchiarini, C. : 'Look, I can speak correctly' : learning vocabulary and pronunciation through websites equipped with automatic speech recognition technology. Computer Assisted Language Learning **0**(0) (2022)
7. Bruton, A. : The Vocabulary Knowledge Scale : A Critical Analysis. Language Assessment Quarterly **6**(4), 288–297 (2009)
8. Bueno-Alastuey, M.C., Nemeth, K. : Quizlet and podcasts : effects on vocabulary acquisition. Computer Assisted Language Learning **0**(0) (2020)
9. Chen, C.M., Chen, L.C., Yang, S.M. : An English vocabulary learning app with self-regulated learning mechanism to improve learning performance and motivation. Computer Assisted Language Learning **32**(3), 237–260 (2019)
10. Chen, C.M., Liu, H., Huang, H.B. : Effects of a mobile game-based English vocabulary learning app on learners' perceptions and learning performance : A case study of Taiwanese EFL learners. ReCALL **31**(2), 170–188 (2019)
11. Chukharev-Hudilainen, E., Klepikova, T.A. : The effectiveness of computer-based spaced repetition in foreign language vocabulary instruction : a double-blind study. CALICO Journal **33**(3), 334–354 (2016)
12. Fang, W.C., Yeh, H.C., Luo, B.R., Chen, N.S. : Effects of mobile-supported task-based language teaching on EFL students' linguistic achievement and conversational interaction. ReCALL **33**(1), 71–87 (2021)

13. Hao, T., Wang, Z., Ardasheva, Y. : Technology-Assisted Vocabulary Learning for EFL Learners : A Meta-Analysis. Journal of Research on Educational Effectiveness **14**(3), 645–667 (2021)
14. Hava, K. : Exploring the role of digital storytelling in student motivation and satisfaction in EFL education. Computer Assisted Language Learning **34**(7) (2021)
15. Hsu, T.C. : Learning English with Augmented Reality : Do learning styles matter ? Computers & Education **106**, 137–149 (2017)
16. Hwang, G.J., Wang, S.Y. : Single loop or double loop learning : English vocabulary learning performance and behavior of students in situated computer games with different guiding strategies. Computers & Education **102**, 188–201 (2016)
17. Jeon, J. : Chatbot-assisted dynamic assessment (CA-DA) for L2 vocabulary learning and diagnosis. Computer Assisted Language Learning **0**(0), 1–27 (2021)
18. Khezrlou, S., Ellis, R., Sadeghi, K. : Effects of computer-assisted glosses on EFL learners' vocabulary acquisition and reading comprehension in three learning conditions. System **65**, 104–116 (2017)
19. Klimova, B. : Evaluating Impact of Mobile Applications on EFL University Learners' Vocabulary Learning – A Review Study. Procedia Computer Science **184**, 859–864 (2021)
20. Kosmas, P., Zaphiris, P. : Words in action : investigating students' language acquisition and emotional performance through embodied learning. Innovation in Language Learning and Teaching **14**(4) (2020)
21. Köse, T., ÿimen, E., Mede, E. : Perceptions of EFL Learners about Using an Online Tool for Vocabulary Learning in EFL Classrooms : A Pilot Project in Turkey. Procedia - Social and Behavioral Sciences **232**, 362–372 (2016)
22. Larchen Costuchen, A., Darling, S., Uytman, C. : Augmented reality and visuospatial bootstrapping for second-language vocabulary recall. Innovation in Language Learning and Teaching **15**(4) (2021)
23. Li, R., Meng, Z., Tian, M., Zhang, Z., Xiao, W. : Modelling Chinese EFL learners' flow experiences in digital game-based vocabulary learning : the roles of learner and contextual factors. Computer Assisted Language Learning **34**(4), 483–505 (2021)
24. Li, Y., Hafner, C.A. : Mobile-assisted vocabulary learning : Investigating receptive and productive vocabulary knowledge of Chinese EFL learners. ReCALL **34**(1), 66–80 (2022)
25. Lin, J.J., Lin, H. : Mobile-assisted ESL/EFL vocabulary learning : a systematic review and meta-analysis. Computer Assisted Language Learning **32**(8), 878–919 (2019)
26. Lin, V., Yeh, H.C., Huang, H.H., Chen, N.S. : Enhancing EFL vocabulary learning with multimodal cues supported by an educational robot and an IoT-Based 3D book. System **104**, 102691 (2022)
27. Liu, C.C., Wang, P.C., Tai, S.J.D. : An analysis of student engagement patterns in language learning facilitated by Web 2.0 technologies. ReCALL **28**(2), 104–122 (2016)
28. Liu, P.L. : Mobile english vocabulary learning based on concept mapping strategy. Language Learning p. 14 (2016)
29. Mehdiyev, E., Uğurlu, C., Usta, H. : Validity and reliability study : Motivation scale in english language teaching. The Journal of Academic Social Science Studies **54**, 21–37 (2017)

30. Mirzaei, A., Domakani, M.R., Rahimi, S. : Computerized lexis-based instruction in EFL classrooms : Using multi-purpose LexisBOARD to teach L2 vocabulary. ReCALL **28**(1), 22–43 (2016)
31. Mirzaei, S., Lewis, T., Wyra, M., Wilkinson, B. : Usability Testing of VLASTWA : A Vocabulary and Strategy Teaching Web App. In : 32nd Australian Conference on Human-Computer Interaction. pp. 614–621. OzCHI '20, New York, NY, USA (2020)
32. Nation, I.S.P. : Teaching and learning vocabulary. Boston, Mass, nachdr. edn. (1999)
33. Pintrich, P.R., Others, A. : A Manual for the Use of the Motivated Strategies for Learning Questionnaire (MSLQ). Tech. rep. (1991)
34. Sato, T., Murase, F., Burden, T. : An Empirical Study on Vocabulary Recall and Learner Autonomy through Mobile ?Assisted Language Learning in Blended Learning Settings. CALICO Journal **37**(3), 254–276 (2020)
35. Song, Y., Wen, Y., Yang, Y., Cao, J. : Developing a 'Virtual Go mode' on a mobile app to enhance primary students' vocabulary learning engagement : an exploratory study. Innovation in Language Learning and Teaching **0**(0) (2022)
36. Soyoof, A., Reynolds, B.L., Shadiev, R., Vazquez-Calvo, B. : A mixed-methods study of the incidental acquisition of foreign language vocabulary and healthcare knowledge through serious game play. Computer Assisted Language Learning **0**(0), 1–34 (2022)
37. Stockwell, G., Liu, Y.C. : Engaging in mobile phone-based activities for learning vocabulary : An investigation in Japan and Taiwan. CALICO Journal **32**(2), 299–322 (2015)
38. Tai, T.Y., Chen, H.H.J., Todd, G. : The impact of a virtual reality app on adolescent EFL learners' vocabulary learning. Computer Assisted Language Learning **35**(4), 892–917 (2022)
39. Tsay, C.H.H., Kofinas, A.K., Trivedi, S.K., Yang, Y. : Overcoming the novelty effect in online gamified learning systems : An empirical evaluation of student engagement and performance. Journal of Computer Assisted Learning **36**(2), 128–146 (2020)
40. Wang, F., Zhang, R., Zou, D., Au, O., Xie, H., Wong, L. : A review of vocabulary learning applications : From the aspects of cognitive approaches, multimedia input, learning materials, and game elements. Knowledge Management and E-Learning **13**(3), 250–272 (2021)
41. Webster, J., Watson, R.T. : Analyzing the past to prepare for the future : writing a litterature review p. 11 (2002)
42. Wu, Q. : Designing a smartphone app to teach English (L2) vocabulary. Computers & Education **85**, 170–179 (2015)
43. Yang, Q.F., Chang, S.C., Hwang, G.J., Zou, D. : Balancing cognitive complexity and gaming level : Effects of a cognitive complexity-based competition game on EFL students' English vocabulary learning performance, anxiety and behaviors. Computers & Education **148**, 103808 (2020)
44. Yihong, G., Yuan, Z., Ying, C., Yan, Z. : Relationship Between English Learning Motivation Types and Self-Identity Changes Among Chinese Students. TESOL Quarterly **41**(1) (2007)

45. Yu, A., Trainin, G. : A meta-analysis examining technology-assisted L2 vocabulary learning. ReCALL **34**(2), 235–252 (2022)
46. Yüksel, H.G., Mercanoğlu, H.G., Yılmaz, M.B. : Digital flashcards vs. wordlists for learning technical vocabulary. Computer Assisted Language Learning **0**(0) (2020)
47. Zhonggen, Y. : Differences in serious game-aided and traditional English vocabulary acquisition. Computers & Education **127**, 214–232 (2018)
48. Zou, D., Huang, Y., Xie, H. : Digital game-based vocabulary learning : where are we and where are we going ? Computer Assisted Language Learning **34**(5-6), 751–777 (2021)
49. Çakmak, F., Erçetin, G. : Effects of gloss type on text recall and incidental vocabulary learning in mobile-assisted L2 listening. ReCALL **30**(1), 24–47 (2018)

Learning Analytics et recherche scientifique au prisme du droit des données personnelles : contribution d'un projet de recherche en Éducation aux Médias et à l'Information

Tanja Petelin[1], Hassina El Kechaï[2]

[1] Université de Poitiers, CECOJI, UR- 21665 Poitiers, France
tanja.petelin@ univ-poitiers.fr
[2] Université de Poitiers, TECHNE, UR- 20297 Poitiers, France
hassina.el.kechai@ univ-poitiers.fr

Résumé. Le travail décrit dans cet article s'inscrit dans le contexte des *Learning Analytics* et porte sur la mise en œuvre des orientations majeures du RGPD dans le cadre des recherches en EIAH ayant recours à la collecte de données à caractère personnel. Il décrit un cas d'usage dans le contexte de la conception d'un dispositif dans le domaine de l'EMI (Éducation aux Médias et à l'Information) en contexte scolaire pour un public de lycéens. Ce dispositif porte sur la collecte de leurs données d'interaction issues de l'utilisation de leurs Smartphones et leur restitue leurs pratiques numériques sous forme de tableaux de bord. Nous nous focalisons dans cet article sur la mise en conformité du RGPD du dispositif que nous illustrons par quelques exemples.

Mots-clés : RGPD, donnée à caractère personnel, Éducation aux Médias et à l'Information, *Learning Analytics*

Abstract. The article provides a proposal of guidelines describing the main principles of GDPR compliance in the context of EIAH research that involves the collection of personal data. The applicative context of this guide is in the design of a device in the field of EMI (Media and Information Education) in a school context for a high school audience and deals with the collection of their interaction data from smartphones. We provide some examples of the implementation of GDPR in this context.

Keywords: GDPR, personal data, Media and Information Education, Learning Analytics.

1 Introduction

Le Règlement Général sur la Protection des Données à caractère personnel "RGPD" [12] est entré en vigueur le 25 mai 2018. Ce texte, d'application directe dans tous les États membres de l'Union européenne, laisse néanmoins à ces derniers une marge de

manœuvre sur certains points. En France, la loi informatique et libertés est ainsi maintenue : en complément du règlement, elle intègre notamment des dispositions d'adaptation du RGPD au contexte national [10]. Le RGPD concerne tout acteur privé ou public établi sur le territoire de l'Union européenne, que le traitement ait lieu ou non dans l'Union, ainsi que ceux non établis dans l'Union dès lors que les activités de traitement ciblent des personnes concernées qui se trouvent sur le territoire de l'Union (RGPD, art. 3). Les travaux de recherche en EIAH ont de plus en plus recours à la collecte et à l'analyse de données d'apprentissage. Ces travaux ont rendu populaire le champ disciplinaire des *Learning Analytics*, ou « analyse de l'apprentissage » qui vise à recueillir les données liées à un dispositif d'apprentissage, à les analyser puis à présenter les résultats de l'analyse sous forme de rapports par exemple ou de tableaux de bord [5]. Pour notre part, nous collectons les données pour réaliser des analyses d'usages et la restitution de ces analyses peut s'adresser à des apprenants pour les mettre dans une démarche réflexive sur leurs pratiques ou à des enseignants pour réguler des pratiques numériques. Cela peut également servir à des institutions pour mesurer la pertinence d'une politique d'équipement numérique, pour produire des connaissances sur les pratiques numériques des élèves ou encore pour accompagner les transformations pédagogiques des enseignants [11]. Nous nous inscrivons donc dans l'analyse des comportements en termes d'utilisation du numérique et en aucun cas sur l'analyse de la performance en termes d'apprentissage. Tous ces travaux sont tenus de s'assurer que le traitement des données personnelles collectées au travers de différents dispositifs de *Learning Analytics* soit conforme au RGPD. Traiter les données conformément au RGPD signifie qu'il faut encadrer les données tout le long de leur cycle de vie à savoir : leur collecte, leur conservation, leur exploitation, leur modification, leur communication, leur archivage, jusqu'à leur destruction ou anonymisation. C'est ainsi que le RGPD doit non seulement accompagner tout le processus de conception et de développement de dispositifs utilisant les *Learning Analytics*, mais peut aller jusqu'à les façonner en termes d'interfaces notamment.

Alors que le RGPD est parfois perçu comme une contrainte[1], pouvant limiter la mise en œuvre des objectifs de la recherche, son analyse plus approfondie révèle le contraire. En effet, le RGPD vise à accorder aux personnes physiques un contrôle sur leurs données personnelles et impose aux responsables de traitement une démarche de conformité, qui doit être rigoureusement respectée et documentée (RGPD, art. 24). Pourtant, loin de constituer un obstacle, chacune des étapes offre une certaine souplesse. Les besoins particuliers de la recherche scientifique donnent d'ailleurs lieu à quelques dérogations[2], tout en rappelant, dans ce cadre, l'obligation de mettre en place des garanties appropriées pour les droits et libertés des personnes (RGPD, art. 89). Le RGPD constitue ainsi un cadre obligatoire, basé sur un socle de principes directeurs, la responsabilisation des acteurs et le respect des droits des personnes concernées : sa mise en œuvre concrète dans le domaine des EIAH peut constituer une opportunité pour le développement d'une recherche respectueuse des droits des personnes sans sacrifier les objectifs scientifiques.

[1] Complexité du travail avec les DPO, difficulté de trouver des solutions ciblées et souples dans les articles du RGPD, changement de posture du public cible dû à la connaissance des finalités de traitement de la recherche imposée par le RGPD sur le principe de transparence...

[2] V. notamment art. 5.1 b) (limitation des finalités) et e) (limitation de la durée de conservation).

Dans cet article, nous traitons de la problématique de la prise en compte du RGPD dans la conception de dispositifs de Learning Analytics et de sa mise en œuvre dans le domaine des EIAH. Pour répondre à cette problématique et après la présentation d'un état de l'art qui traite de la protection des données personnelles et des règles de conduite dictées par le RGPD, nous présentons les orientations importantes à prendre en compte pour s'assurer que les dispositifs soient conformes aux règles établies par le RGPD. Nous proposons ensuite quelques exemples de mise en œuvre de ces orientations dans le cadre du processus de conception et de développement d'un dispositif en Éducation aux Médias et à l'Information (EMI) dans le contexte scolaire.

2 État de l'art

2.1 Protection des données personnelles

Le RGPD définit une donnée personnelle comme « toute information se rapportant à une personne physique identifiée ou identifiable » (art. 4.1). Ces informations peuvent être des identifiants, tels qu'un nom, un numéro d'identification, les données de localisation, un identifiant en ligne, ou encore des informations plus spécifiques, propres à son identité physique, physiologique, génétique, psychique, économique, culturelle ou sociale. Toutes les données personnelles ne sont pas à mettre au même niveau. En effet, certaines peuvent être considérées comme "sensibles" et peuvent être soumises à un régime juridique de protection renforcée, basée sur le principe d'interdiction de traitement, assorti d'une liste des exceptions. Les données dites sensibles peuvent être par exemple : l'origine raciale ou ethnique, l'opinion politique ou religieuse, l'appartenance syndicale, les données relatives à la santé, à la vie ou à l'orientation sexuelle (RGPD, art. 9).

Le RGPD s'applique dès lors qu'un dispositif traite des données personnelles, ce dernier terme étant défini comme « toute opération ou tout ensemble d'opérations effectuées ou non à l'aide de procédés automatisés [...] telles que la collecte, l'enregistrement, l'organisation, la structuration, la conservation, l'adaptation ou la modification, l'extraction, la consultation, l'utilisation, la communication par transmission, la diffusion ou toute autre forme de mise à disposition, le rapprochement ou l'interconnexion, la limitation, l'effacement ou la destruction » (RGPD, art. 4.2).

Le traitement des données personnelles doit être distingué du traitement des données anonymes ou anonymisées : ces dernières ne rentrent pas dans le champ d'application de la protection des données personnelles. La technique d'anonymisation doit cependant correspondre aux exigences strictes de la CNIL et « empêcher toutes les parties d'isoler un individu dans un ensemble de données, de relier entre eux deux enregistrements dans un ensemble de données (ou dans deux ensembles de données séparés) et de déduire des informations de cet ensemble de données » [4] [7]. L'anonymisation est ainsi incompatible avec les traitements dont la finalité est le suivi du comportement ou de l'activité d'une personne [4]. En effet, cette technique ne doit pas être confondue avec la pseudonymisation, consistant à remplacer le traitement des informations directement identifiables (nom, adresse IP...) par des données non directement identifiables (p. ex. un numéro aléatoire). Dans ce second cas, il est possible de retrouver l'identité

de la personne si l'on dispose d'informations supplémentaires : la législation sur la protection des données personnelles s'applique.

2.2 RGPD et applications de *Learning Analytics* : règles de conduite

Lorsque l'on conçoit ou l'on utilise des dispositifs de *Learning Analytics*, le principe de *Privacy by design* impose la mise en œuvre des principes relatifs à la protection des données tant au moment de la détermination des moyens du traitement qu'au moment du traitement lui-même (RGPD, art. 25.1). Les chercheurs en EIAH qui conçoivent des dispositifs intégrant des *Learning Analytics* doivent donc prendre en compte les principes et règles de protection des données personnelles dès la phase de conception des dispositifs de collecte et d'analyse des données. Selon l'article 5 du RGPD, le traitement des données doit respecter :

- **Les principes de licéité, de loyauté et de transparence :** En premier lieu, ceci impose de disposer d'un *fondement de licéité* qui, en matière de *Learning Analytics*, sera souvent le consentement de la personne concernée. Toutefois, certaines applications de *Learning Analytics* pourraient être basées sur d'autres fondements, cités dans l'article 6 du RGPD, tels que l'intérêt légitime du responsable de traitement ou d'un tiers (à moins que ne prévalent les intérêts ou les libertés et droits fondamentaux de la personne concernée) ou, le cas échéant, mission d'intérêt public (une telle mission doit néanmoins être définie par le droit national ou européen – elle ne se présume pas) voire même sur une obligation légale si, dans l'avenir, le législateur impose à certains acteurs de procéder à l'analyse des données d'apprentissage. Quant à la transparence et la loyauté, leur mise en œuvre repose sur une démarche active de communication, compréhensible et aisément accessible, de toute information pertinente, conformément aux exigences du RGPD (art. 12 à 14). Par ailleurs, la personne concernée ne doit jamais être prise au dépourvu à un stade ultérieur quant à la façon dont ses données à caractère personnel ont été utilisées [9].
- **Le principe de limitation des finalités du traitement :** Les données doivent être collectées à des fins préalablement déterminées, explicites et légitimes. Elles ne doivent pas être traitées ultérieurement d'une manière incompatible avec ces finalités. Toutefois, le RGPD prévoit que le traitement ultérieur à des fins de recherche scientifique n'est pas considéré, conformément à l'article 89, paragraphe 1, comme incompatible avec les finalités initiales.
- **Le principe de minimisation des données (ou de nécessité)** : Seules les données adéquates, pertinentes et limitées à ce qui est nécessaire au regard des finalités peuvent être collectées.
- **Le principe d'exactitude** : Les données doivent être exactes, en cas d'inexactitude, les données doivent être effacées ou rectifiées.
- **Le principe de limitation de la conservation** : Les données ne doivent pas être conservées pour une durée supérieure à ce qui est nécessaire au regard de la finalité du traitement.

- **Les principes d'intégrité et de confidentialité :** Les données doivent être collectées et conservées selon des procédés garantissant la protection contre les intrusions, la perte et la destruction des données.

L'obligation de mise en conformité s'applique en premier lieu au « responsable du traitement » défini comme « la personne physique ou morale, l'autorité publique, le service ou un autre organisme qui [...] détermine les finalités et les moyens du traitement ». Dans le cadre des recherches universitaires, les laboratoires ne disposant pas de personnalité juridique, cette qualité revient formellement à l'université dont relèvent les chercheurs. Lorsque les finalités et les moyens du traitement sont définis conjointement par plusieurs responsables de traitement (par exemple dans le cadre d'une collaboration entre plusieurs laboratoires ou avec des structures privées), ceux-ci ont la qualité de responsables conjoints, qui doivent définir de manière transparente leurs obligations respectives au regard du RGPD (art. 26). Enfin, lorsqu'un acteur procède à des traitements des données personnelles en tant que sous-traitant, c'est-à-dire pour le compte d'un responsable de traitement, le RGPD impose des obligations spécifiques (art. 28). Le statut de chacun des intervenants au regard du RGPD dépendra donc de la question de savoir s'il participe ou non à la détermination des finalités et des moyens de traitements.

3 Accompagner l'intégration du RGPD dans la conception de dispositifs de *Learning Analytics* en EIAH

3.1 Démarche de conformité : orientations générales

Malgré le travail d'accompagnement effectué par la CNIL, le RGPD suscite encore de nombreuses interrogations et continue de poser des difficultés dans les projets de conception et d'utilisation de dispositifs de *Learning Analytics* soit par manque de ressources, soit par méconnaissance de la réglementation. Le RGPD a en effet bousculé la manière de concevoir et de développer des dispositifs intégrant des *Learning Analytics*. Bien que des efforts sensibles soient effectués, les chercheurs ont le sentiment qu'il est difficile de s'approcher d'un niveau de conformité élevé sans impacter les objectifs de leurs recherches. Or ce sentiment peut résulter d'un manque de maturité des travaux de recherche en EIAH en termes de mise en conformité : les mécanismes permettant de concilier la protection des données personnelles avec les finalités de la recherche scientifique ne sont pas encore suffisamment mis en œuvre et des incertitudes persistent. Afin d'aider les chercheurs en EIAH à bien intégrer les principes de conformité dans les dispositifs de *Learning Analytics*, nous présentons ci-dessous un guide des principales orientations à prendre en compte et à contextualiser.

Première étape : Dans le cadre d'une recherche menée au sein d'une université ou d'un organisme public, une des premières étapes consiste à se rapprocher du délégué à la protection des données (DPO). Ce dernier est le coordinateur de la mise en conformité au sein de la structure et gère notamment le registre des activités de traitement de

celle-ci, auquel devra être inscrit le traitement envisagé en *Learning Analytics*. Il veillera à ce que le traitement soit conforme au RGPD et sollicitera toute information permettant de démontrer cette conformité, souvent à travers un questionnaire à remplir.

Deuxième étape : Après l'identification des personnes chargées de la réalisation du traitement et la détermination de ce dernier, il faudra expliquer, conformément au *principe de limitation des finalités* de traitement (cf fig.1, point 1), la raison d'être de celui-ci. En effet, comme vu précédemment, les données personnelles ne peuvent être traitées que pour des finalités déterminées, explicites et légitimes. La finalité ainsi définie par les chercheurs en EIAH aura un impact sur la détermination des données qui pourront être traitées, sur la durée de leur conservation, ainsi que sur la définition du fondement de licéité du traitement. Le rôle central de la détermination des finalités et son interaction avec la mise en œuvre des autres principes sont illustrées dans la figure n° 1.

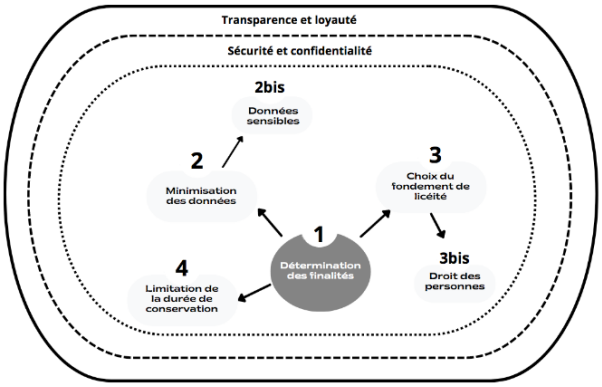


Fig. 1. Les orientations de mise en œuvre des principes du RGPD en EIAH.

Premièrement, selon le *principe de minimisation des données* (cf fig.1, point 2), la finalité justifiera l'étendu des données collectées : avant la mise en œuvre de la collecte, les chercheurs devront ainsi apprécier l'adéquation des données collectées par rapport aux finalités de la recherche et éviter une collecte excessive des données, allant au-delà de ce qui est nécessaire. Par ailleurs, la finalité de la recherche peut parfois imposer la collecte des données sensibles (cf fig.1, point 2bis), telles que définies dans l'article 9 du RGPD et citées plus haut (opinions politiques, religion, santé...). L'identification de ces données n'est pas toujours aisée, car le caractère sensible d'une donnée est parfois indirect. Or, selon un récent arrêt de la Cour de justice de l'Union européenne [2], le régime des données sensibles semble s'y appliquer. Nous pouvons notamment citer les données de géolocalisation qui peuvent par exemple révéler la fréquentation d'un lieu de culte et donc révéler la religion réelle ou supposée d'une personne. Pourtant, ce type de données a parfois un intérêt pour la recherche. Ainsi, dans le projet d'EMI, présenté en section 3.2, les données de géolocalisation permettent d'analyser l'itinérance des pratiques numériques des lycéens et leur circulation dans les sphères personnelle et scolaire, identifiées par les lieux fréquentés. Le traitement des données sensibles étant en principe interdit, une telle conception large des données sensibles peut remettre en

question les pratiques des chercheurs. Toutefois, des exceptions autorisant le traitement de ces données existent (par exemple le consentement explicite de la personne ou, plus spécialement pour la recherche publique, des motifs d'intérêt public important exigeant néanmoins un avis motivé et publié de la CNIL) et peuvent être mobilisées dans le cadre d'une recherche en EIAH. Il convient de rappeler que les conditions de ces exceptions doivent être réunies en plus des conditions du *fondement de licéité* appliqué (V. ci-dessous).

Deuxièmement, la finalité de la collecte aura également un impact sur la détermination du *fondement de licéité* (cf fig.1, point 3). En plus du consentement de la personne concernée, l'article 6.1 du RGPD prévoit cinq autres fondements : si le traitement est nécessaire à la réalisation d'une des fins prévues dans ces cinq hypothèses, ce fondement peut être retenu (nous avons déjà évoqué, en matière de *Learning Analytics*, l'exécution d'une mission d'intérêt public ou la poursuite d'un intérêt légitime) ; dans le cas contraire, le seul fondement rendant licite le traitement est le consentement de la personne concernée (voire de son représentant légal en cas de mineurs). Le choix du fondement de licéité aura un impact sur les droits que le RGPD (art. 15 à 21) reconnaît aux personnes concernées (cf fig.1, point 3bis). Ainsi, lorsque le traitement est fondé sur le consentement, la personne concernée a le droit de le retirer à tout moment, de manière simple et équivalente à celle utilisée pour le recueillir. À partir du retrait du consentement, en principe, tout traitement des données doit cesser et ces dernières doivent être effacées. Pourtant, la suppression des données risque parfois de compromettre les objectifs de la recherche. Cette considération a donné lieu à une exception en matière de recherche scientifique, permettant de déroger à l'obligation d'effacement lorsque ceci peut rendre impossible ou compromettre gravement la réalisation des objectifs du traitement. Alors que le champ d'application concret de cette exception reste à déterminer, lorsque la recherche peut utiliser les données anonymisées, cette solution est à privilégier. Toutefois, l'anonymisation constitue aussi un traitement des données personnelles, qui doit respecter les exigences du RGPD et être basé sur un des fondements de licéité [7]. Le consentement pour le traitement initial ayant été retiré, il serait éventuellement possible, nous semble-t-il, de baser l'anonymisation sur l'intérêt légitime du responsable du traitement (poursuite de la recherche) tout en informant les personnes concernées (transparence). Ce problème, propre aux traitements fondés sur le consentement, ne se pose pas pour les traitements initialement fondés sur la mission d'intérêt public ou sur l'intérêt légitime. Or, bien que l'article 21 du RGPD prévoie un droit d'opposition dans ces deux hypothèses, ce droit n'est pas absolu. En effet, selon cet article, le traitement peut être poursuivi s'il « *existe des motifs légitimes et impérieux pour le traitement qui prévalent sur les intérêts et les droits et libertés de la personne concernée* ». Plus spécifiquement, dans le domaine de la recherche scientifique, il est possible d'y déroger lorsque l'exercice de ce droit risque de rendre impossible ou d'entraver sérieusement la réalisation des finalités de la recherche[3]. Ces deux fondements de licéité peuvent donc offrir plus de sécurité aux chercheurs quant à la possibilité de mener à terme la recherche nécessitant des données personnelles. Toutefois, leurs conditions ne sont pas toujours réunies, la mission d'intérêt public devant être explicite-

[3] RGPD, art. 89.2 ; LIL, art. 78 ; Décr. n° 2019-536 du 29 mai 2019, art. 116

ment prévue par le droit et l'intérêt légitime ne pouvant justifier le traitement que lorsque *« ne prévalent les intérêts ou les libertés et droits fondamentaux de la personne concernée qui exigent une protection des données à caractère personnel, notamment lorsque la personne concernée est un enfant* » (balance des intérêts).

Enfin, selon le principe de *limitation de la conservation* des données (cf fig.1, point 4), la durée de la conservation ne doit pas excéder ce qui est nécessaire au regard des finalités pour lesquelles elles sont traitées. Dans une démarche de conformité, il conviendra donc de déterminer une telle durée conforme aux finalités de la recherche et veiller à ce que celle-ci soit respectée : à l'expiration de cette durée, les données doivent être supprimées ou anonymisées. Pour répondre aux besoins particuliers de la recherche scientifique (notamment l'exigence de la reproductibilité de la recherche), le RGPD prévoit la possibilité de déroger à ce principe, sous condition de mettre en œuvre les mesures techniques et organisationnelles appropriées pour garantir la sécurité et la confidentialité des données. Toutefois, lorsque les mêmes finalités peuvent être atteintes par des données anonymisées, cette voie est à privilégier[4]. En parallèle (cf. fig1, cercle 2), il faudra toujours garantir la sécurité des données à caractère personnel à l'aide de mesures techniques ou organisationnelles appropriées (*intégrité et confidentialité*). À titre non limitatif, le RGPD (art. 32) cite plusieurs techniques permettant d'assurer la sécurité du traitement, notamment la pseudonymisation[5], fortement encouragée par le RGPD (y compris dans le régime des garanties et dérogations applicables aux traitements à des fins de recherche scientifique[6]). De plus, conformément au *principe de loyauté et de transparence*, le traitement doit être transparent à l'égard des personnes concernées (cf fig.1, cercle 3). Les articles 12 à 14 du RGPD imposent, respectivement, une certaine qualité de communication des informations, qui doit être *« concise, transparente, compréhensible et aisément accessible, en des termes clairs et simples, en particulier pour toute information destinée spécifiquement à un enfant* », et une liste des informations qui doivent être communiquées.

En principe, depuis l'entrée en vigueur du RGPD, la mise en œuvre de la protection des données personnelles, telle que décrite ci-dessus, relève des responsables de traitement : le rôle de la CNIL consiste, le cas échéant, à contrôler cette conformité. Toutefois, certains traitements en matière de *Learning Analytics* peuvent présenter des risques élevés pour les droits et libertés des personnes physiques (surveillance systématique de leur comportement en ligne, données des mineurs...). La CNIL définit ainsi, sur la base de l'article 35 du RGPD, une liste des types d'opérations de traitement pour lesquelles il est nécessaire de réaliser une analyse d'impact relative à la protection des données (AIPD) [3], qui complète la liste de 9 critères issus des lignes directrices

[4] V. RGPD, art. 89.1

[5] Celle-ci est définie comme « le traitement de données à caractère personnel de telle façon que celles-ci ne puissent plus être attribuées à une personne concernée précise sans avoir recours à des informations supplémentaires, pour autant que ces informations supplémentaires soient conservées séparément et soumises à des mesures techniques et organisationnelles afin de garantir que les données à caractère personnel ne sont pas attribuées à une personne physique identifiée ou identifiable est d'ailleurs » (RGPD, art. 4.5)

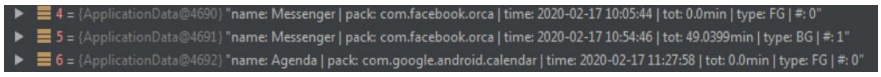
[6] RGPD, art. 89

du G29 [8]. Il convient alors, à l'aide d'une méthodologie adaptée[7] et en collaboration avec le DPO, d'apprécier la gravité et la vraisemblance des risques, ainsi que les mesures mises en place pour les prévenir. S'il apparait que le niveau de risque résiduel reste élevé malgré ces mesures, la CNIL doit être consultée avant la mise en œuvre du traitement.

3.2 Mise en œuvre dans un projet d'EMI en contexte scolaire

La mise en œuvre que nous décrivons est réalisée dans le cadre d'un projet de recherche qui s'inscrit dans le domaine de l'EMI (Éducation aux Médias et à l'Information) dans le contexte scolaire. La contribution par la recherche de ce projet porte sur l'éducation des adolescents à l'usage raisonné du smartphone pour réguler leur temps de sommeil et l'allocation de leur attention sur le temps scolaire. Le smartphone étant un équipement personnel, toujours disponible, nomade et connecté, il est propice aux usages personnels quotidiens, ce qui impacte le sommeil et l'attention des adolescents quand son utilisation n'est pas raisonnée. Le public ciblé est un public de lycéens entre 15 et 18 ans qui sont les plus exposés aux risques numériques.

Afin d'inciter les lycéens à un usage raisonné du smartphone, nous avons développé une application mobile dans le cadre de l'enseignement de l'EMI qui leur restitue leurs pratiques numériques [1] sous forme de tableaux de bord. La restitution est rendue possible grâce à la collecte et à l'analyse de leurs données d'interaction issues de l'utilisation du smartphone. La collecte de données est réalisée grâce à une application mobile qui « écoute » toutes les actions de l'utilisateur et transmet les données d'interaction sur un serveur sécurisé. Cette application est constituée de deux couches : une couche Application mobile à installer pour collecter les données d'interaction des adolescents, une couche Serveur caractérisée par une base de données qui stocke les données collectées et une API qui fait office de module de transmission des données entre l'application mobile et la base de données. Plusieurs données sont collectées, par exemple : les actions de marche/veille du smartphone, la consommation de données par application, les applications utilisées, les interactions avec les applications, les URLs visitées, la localisation, les notifications et réactions utilisateurs à ces notifications… À titre d'illustration, nous présentons dans la figure 2 un exemple de données brutes décrivant les interactions avec les applications ainsi que son format. Il s'agit de collecter des données décrivant les applications ouvertes par l'utilisateur et comment l'utilisateur interagit avec elles à travers l'attribut "action" : FG (foreground) quand une application passe en premier plan suite à l'interaction de l'utilisateur et BG (background) quand l'application passe en arrière-plan. Nous retrouvons également le nom de l'application, l'horodatage ainsi que la durée de l'interaction.



[7] La CNIL met à dispositions du public un logiciel open source (PIA : Privacy Impact Assessment), afin de faciliter la conduite et la formalisation d'AIPD.

Fig. 2. Exemples de données brutes décrivant les interactions avec les applications : Format données = {(application, package, horodatage, durée, action (FG | BG))

Dans cette section, nous présentons la mise en œuvre des orientations décrites dans la section 3.1 dans le projet d'EMI. Tous les principes évoqués et ces orientations ont été mis en œuvre mais nous ne pouvons pas tous les décrire. Nous nous focalisons sur certains critères ayant donné lieu à des fonctionnalités dans l'application d'EMI que nous pouvons illustrer.

3.2.1 Mise en œuvre du principe de transparence

Celle-ci prend la forme d'une **notice d'information** que nous avons rédigée et qui précise l'ensemble des éléments qui doivent être communiquées selon l'article 13 du RGPD (identification du responsable de traitement, les finalités du traitement, la base légale, les données traitées, les destinataires des données, les modalités d'exercice des droits des personnes, la durée de conservation,...). Conformément à l'article 12 du RGPD, cette notice a été intégrée dans l'application de collecte de telle manière qu'elle puisse être facilement accessible à l'utilisateur tout le temps. Elle se présente dans l'onglet « Données personnelles » de l'application (Figure 3).

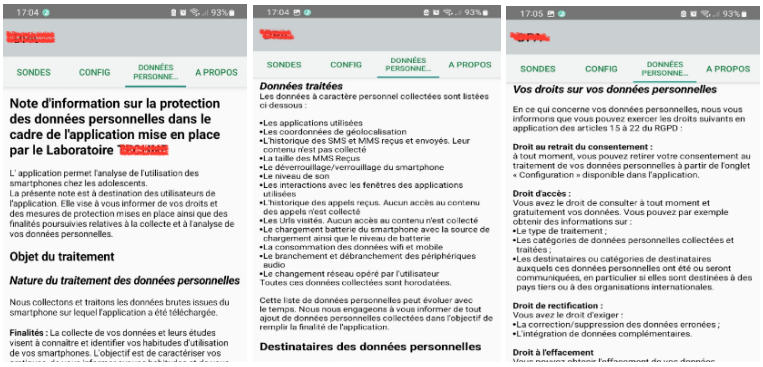


Fig. 3. Quelques extraits de la notice d'information intégrée dans l'application d'EMI

3.2.2 Mise en œuvre du principe de licéité : consentement

Le fondement de licéité retenu dans le cadre de cette recherche est le consentement. L'application de collecte a été façonnée de sorte à permettre l'expression d'une volonté libre, spécifique, *éclairée* et *univoque*, conformément à l'article 4.11 du RGPD [6]. Pour recueillir le consentement, nous avons mis en œuvre une interface listant les données collectées avec une case à cocher indiquant l'accord de l'utilisateur (cf. fig 4 à gauche) qui peut être décochée pour retirer le consentement à n'importe quel moment. Ce consentement est enregistré sous la forme d'un booléen dans une table de la base de données dédiée au consentement. Pour renforcer le caractère spécifique du consentement de l'utilisateur, nous lui avons également donné la possibilité de choisir les données qu'il autorise à collecter. Ces choix sont opérés par un système « d'interrupteurs » que l'utilisateur peut actionner ou non comme illustré dans la figure 4 à droite. Les

interrupteurs verts correspondent aux données pour lesquelles le consentement est accordé contrairement à celles en rouge.

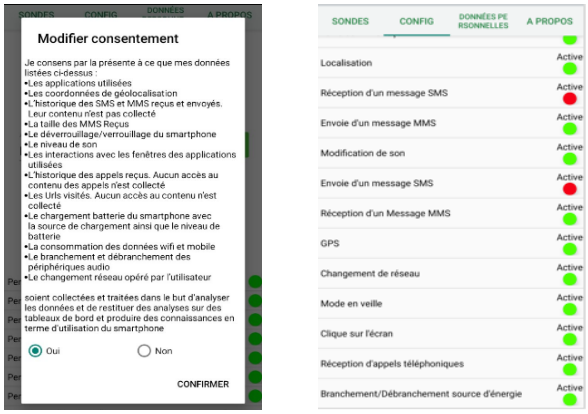


Fig. 4. Le consentement et son retrait (à gauche), sa personnalisation (à droite)

3.2.3 Mise en œuvre du principe de sécurité

Contrairement aux données anonymes, les données personnelles « pseudonymisées » continuent de caractériser une personne [7] en la désignant par un code plutôt que par un nom, comme vu précédemment. Dans notre contexte de suivi des pratiques numériques dans le temps, l'anonymisation n'est pas possible : nous optons donc pour des mesures techniques (script informatique) consistant à pseudonymiser les données de l'utilisateur et à lever cette pseudonymisation pour la restitution dans les tableaux de bord mais de manière totalement opaque et sans que l'on puisse identifier la personne dont les données sont collectées. Dans la figure 5, à droite est présenté un extrait les données de la table "usage des applications" qui décrit les interactions de l'utilisateur. Dans la structure de la table, la pseudonymisation est mise en œuvre à travers l'attribut (device_ref). Il s'agit de la donnée permettant de caractériser l'utilisateur par un code plutôt que par son nom ce qui ne permet pas de le rendre identifiable tout en préservant la possibilité de lui restituer l'analyse de ses données.

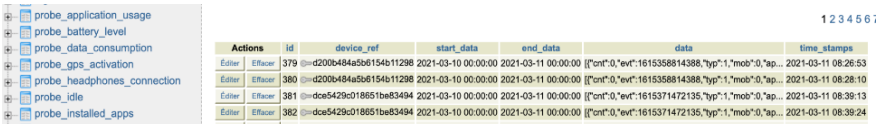


Fig. 5. Base de données et pseudonymisation

4 Conclusion

Dans cet article, nous avons présenté un guide pour la mise en conformité avec le RGPD des travaux de recherche portant sur la conception de dispositifs basés sur la collecte

de données à caractère personnel. Nous avons également donné quelques exemples de mise en œuvre dans le cadre de la conception d'un dispositif d'EMI en contexte scolaire. Cependant, cette mise en œuvre n'est pas présentée dans sa totalité conformément à ce que nous avons réalisé. En effet, il est difficile d'être exhaustif dans cet article car la prise en compte du RGPD dans notre recherche a été un processus long et très documenté. Pour de nombreuses recherches en EIAH, le RGPD apporte de nouvelles difficultés et de nouveaux défis : mise en conformité, acculturation à la donnée, lisibilité de certains métiers comme celui du DPO…Le RGPD et le droit du numérique deviennent des composantes incontournables dans la recherche en EIAH où des dispositifs sont conçus dans une logique de production et d'usages des données. Leur gestion précise, claire et sécurisée est une condition sine qua non pour instaurer un climat de confiance entre chercheurs et utilisateurs de dispositifs d'apprentissage. Les chercheurs en droit du numérique doivent ainsi avoir toute leur place en tant que partenaires des projets de recherche en EIAH contribuant ainsi à enrichir encore plus le caractère pluridisciplinaire du domaine.

Références

1. Aillerie, K.: Pratiques informationnelles informelles des adolescents (14-18 ans) sur le Web. Thèse en Sciences de l'information et de la communication. Université Paris 13 (2011).
2. CJUE, gr. ch., *OT c Vyriausioji tarnybinės etikos komisija*, C184/20 (source jurisprudentielle), 1 août 2022.
3. CNIL, https://www.cnil.fr/fr/RGPD-analyse-impact-protection-des-donnees-aipd [consulté en dernier le 27 décembre 2022].
4. CNIL, Délibération n° 2015-255, aff. JCDecaux. V. aussi, rejetant le recours, CE, 10e et 9e ch., 8 févr. 2017, n° 393714, JCDecaux c/ CNIL 20 (source jurisprudentielle),16 juillet 2015.
5. Dabbebi I, Iksal S., Gilliot JM, May M, Garlatti S: Towards Adaptive Dashboards for Learning Analytic: An Approach for Conceptual Design and implementation in (CSEDU 2017), , Porto, Portugal. pp.120-131, Apr 2017.
6. G29, Lignes directrices sur le consentement au sens du règlement 2016/679, 10 avril 2018.
7. G29, Avis n° 05/2014 sur les techniques d'anonymisation, WP 216, 10 avr. 2014.
8. G29, Lignes directrices concernant l'analyse d'impact relative à la protection des données (AIPD) et la manière de déterminer si le traitement est « susceptible d'engendrer un risque élevé » aux fins du règlement (UE) 2016/679, 4 avr. 2017 (modifiées 4 oct. 2017), WP 248.
9. Les lignes directrices du Groupe de l'article 29 sur la transparence au sens du règlement (UE) 2016/679.
10. Loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés, modifiée par la loi du 20 juin 2018 et par l'ordonnance du 12 décembre 2018 (source légale).
11. Pierrot L., Cerisier JF, El-Kechaï H., Ramírez S, Pottier L: Using a mixed analysis process to identify the students' digital practices. In EC-TEL 2017 Tallinn, Estonia (2017).
12. Règlement (UE) 2016/679 du Parlement Européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE (règlement général sur la protection des données) (source légale).

Favoriser et analyser une auto-récolte des usages du numérique : un enjeu scientifique et méthodologique pour la recherche sur les EIAH

Marion Paggetti[1], Viviane Folcher[1], Emeline Ah-Tchine[1], Anaïs Loizon[1]

[1]Institut Agro Dijon, FoAP, F-21000 Dijon, France
prenom.nom@agrosupdijon.fr

Résumé. Cette contribution prend appui sur les données exploratoires d'un projet de recherche visant à renseigner les usages du « numérique pour apprendre » dans le cadre du recours à de nouveaux artefacts numériques dans l'enseignement supérieur agricole français. Elle se propose de comprendre à quelles conditions il est possible de renseigner des situations d'usage à l'aide de « carnets d'expériences » remplis par les usagers eux-mêmes. Cette première étude offre des pistes à propos des formes d'appropriation des carnets par un groupe-test et des possibilités de transformation d'un vécu en expérience.

Mots-clés : Usages, enseignement supérieur, apprentissage, récolte, carnets.

Abstract. This contribution is based on exploratory data from a research project aimed at documenting the uses of "digital learning" in the context of French academic agricultural education. It aims to understand the conditions under which it is possible to document situations of use with the help of "experiences' notebooks" filled in by the users themselves. This first study offers some clues as to the forms of appropriation of the notebooks by a test group and the possibilities of transforming an experience into an experiment.

Keywords: Uses, academic education, learning, documenting, notebooks.

1 Introduction

L'usage du numérique est omniprésent dans l'enseignement supérieur. Pour les étudiants comme pour les enseignants, il repose la plupart du temps sur une maîtrise de l'outil informatique développée en dehors de tout enseignement formel. Dans un contexte d'élargissement de son utilisation aux situations d'enseignement/apprentissage instituées et non instituées, mais aussi d'enjeux autour de l'appropriation des outils numériques par les usagers, il convient désormais de comprendre les caractéristiques des situations d'usage qui sont significatives pour le sujet ainsi que la nature de ses usages des ressources numériques.

Si l'activité d'apprentissage est réalisée au moyen d'un artefact, l'attention du sujet est centrée sur l'objet de son activité. Dans ce cadre, l'utilisation de l'artefact ainsi que le vécu et l'influence de cette utilisation sur l'activité d'apprentissage appartiennent au registre contextuel de la situation et sont principalement pré-réfléchis [1] ou difficiles à renseigner en tant que tels.

En s'inscrivant dans une perspective de contribution aux processus de conception « pour et dans l'usage » [2] des artefacts numériques pour enseigner et apprendre, cette contribution se propose de comprendre comment recueillir et exploiter des données d'usages recueillies en autonomie par les sujets de la recherche. Elle s'appuie pour cela sur la phase exploratoire d'un projet de recherche conduit dans l'enseignement supérieur agricole français.

Après avoir présenté le contexte et le cadre théorique et conceptuel qui étaient cette contribution empirique, des éléments de résultats préliminaires seront présentés et offriront des perspectives méthodologiques et scientifiques pour envisager le recours à l'auto-récolte des usages du numérique dans la recherche en éducation et formation.

2 Renseigner les transformations numériques dans l'enseignement-apprentissage : le contexte et l'objet de la recherche

Porté par l'alliance Agreenium et financé par l'Agence Nationale de la Recherche, le projet Hercule 4.0© vise à renseigner les transformations qu'implique la pédagogie numérique dans l'enseignement supérieur agricole. Au sein de ses établissements, un état des lieux exhaustif des stratégies d'établissements est notamment réalisé afin de constituer des établissements « démonstrateurs » de l'enseignement supérieur. Il vise à comprendre comment les écoles ont recours au numérique pour apprendre et faire apprendre et ce que cela implique du point de vue de l'activité des enseignants, des apprenants ou encore des politiques institutionnelles.

Un collectif scientifique s'attache à renseigner les usages liés à des modalités pédagogiques utilisant le numérique à des fins d'apprentissage. La généralisation des outils numériques dans le contexte de l'enseignement universitaire donne accès aux outils numériques et informatiques en dehors de tout contexte tutoré. L'emploi et les usages de ces outils échappent aux enseignants comme aux espaces formels d'enseignement-apprentissage et les enjeux de maîtrise et d'appropriation qu'ils recouvrent peinent à être renseignés.

Les sujets sont les seuls à détenir les informations liées aux usages qu'ils font des outils et plateformes à disposition, qu'ils soient sources de difficultés et/ou d'étayage dans l'activité d'apprentissage. Dans cette recherche, le recours aux artefacts numériques par les enseignants et les étudiants est mis en objet et vise à être renseigné « de l'intérieur » par les usagers qui vivent et se « débrouillent » [3] eux-mêmes des situations d'usage.

L'accès fin aux situations d'usage qui composent le quotidien des sujets croise différents types d'enjeux professionnels, scientifiques et méthodologiques. Pour la recherche en éducation et formation par et avec le numérique, l'enjeu consiste à comprendre comment sont utilisés par les apprenants les ressources numériques mises à disposition ou quelles formes d'activités sont à envisager pour accompagner les étudiants dans leurs apprentissages. Pour la communauté scientifique constituée autour des EIAH, l'enjeu consiste à comprendre comment les usagers interagissent effectivement avec les instruments numériques et se les approprient en vue d'apprendre. Enfin, l'approche qui met à contribution les usagers dans la construction de connaissances scientifiques relatives à leurs usages finalisés constitue un enjeu méthodologique de recueil et constitution du corpus de données. Cela demande de construire les conditions et les moyens d'accès aux situations d'usage ainsi qu'à leur possible mode d'exploitation.

3 Etudier les usages du numérique et la mise en mots de l'expérience subjective : le cadre théorique et conceptuel

Cette contribution met en objet l'activité humaine en s'appuyant sur l'idée que tout rapport d'un sujet au monde est médiatisé par des objets, notamment techniques, regroupés sous la dénomination d'artefact [4]. La médiatisation de l'activité réalisée par l'individu dans son milieu physique et social est faite des « pré-construits sociaux qu'il mobilise pour agir qu'il transforme et renouvelle dans ses activités » [5], mais aussi de transformations des objets de l'activité ainsi que d'attributions de fonctions aux artefacts.

En effet, les usages sont toujours constitués d'appropriation ou de potentiels détournements et, en ce sens, ne sont jamais totalement conformes aux emplois initialement prévus lors de la conception. Ils sont toujours mus par un objectif ou un besoin sous-jacent et personnel. Il s'agit de considérer que le recours à un artefact fait systématiquement l'objet de prises de libertés ou de redéfinition des fonctions [4, 5].

La notion d' « usage » employée dans cette contribution renvoie à un enrichissement conceptuel des définitions historiques principalement centrées sur l'appropriation d'artefacts au regard de contraintes et possibles de l'activité [2]. Les usages donnent non seulement à voir les résultats transitoires ou définitifs, fragmentés ou plus globaux de l'activité réalisée, mais aussi ce qui n'a pu survenir dans le réel de l'activité, les difficultés et les empêchements d'agir, et au-delà, les souhaits pour un futur désirable aux formes encore incertaines.

Tandis que l'usage concerne des pratiques habituelles, son étude rend possible celle des évènements ou transformations qui composent le vécu quotidien du sujet. A cet effet, la philosophie pragmatiste issue des travaux de John Dewey propose de considérer les transactions du sujet avec son environnement comme un continuum formé par le « cours habituel de l'activité » [6]. Ponctuellement, le sujet est toutefois confronté à des situations douteuses ou « embrouillées » [3] dans lesquelles il ne dispose pas immédiatement des ressources qui lui permettent d'agir de manière adéquate. Lors de cette rencontre avec une situation perçue comme douteuse, l'émotion émerge chez le sujet et le conduit à sortir du cours habituel de son activité. Un espace est ainsi laissé à une

reconfiguration de l'agir et à la détermination de nouvelles potentialités d'action lors d'une « parenthèse intellective » voire d'une *enquête* [6].

La mise en mots d'une expérience subjective et potentiellement embarrassée permet enfin au sujet de comprendre et conceptualiser les ressorts de cette situation vécue [1] et de l'éventuelle reconfiguration des ressources dont il dispose, éventuellement au moyen d'une enquête sur sa propre activité.

4 Aménager les conditions d'une récolte autonome des usages : démarche méthodologique

La conduite méthodologique du travail est assurée par un collectif constitué de l'équipe de chercheuses ainsi que d'ingénieurs pédagogiques de chaque école de l'alliance Agreenium. A partir du repérage et de la sélection de terrains d'étude privilégiés, des étudiants et enseignants concernés par ces situations pédagogiques sont recrutés sur la base du volontariat et sollicités pour recueillir de manière autonome leurs situations significatives liées à l'usage des outils en situation pédagogique.

Le repérage et la sélection de ces « récolteurs » est suivie d'un temps de formation à la récolte autonome des usages dans des « carnets d'expériences ». En portant la focale sur l'activité située [7], les récolteurs sont encouragés à décrire leur vécu, en renseignant des rubriques : ressenti, description de la situation, difficultés rencontrées, ressources mobilisées. L'instrument « carnet d'expériences » est ainsi structuré afin d'outiller ce geste de récolte de l'expérience singulière. Cette démarche autonome s'appuie ainsi sur l'intention de rendre les participants acteurs de leur propre récit et permet également de recueillir un large panel de situations quotidiennes d'usage.

Une courte phase exploratoire a été conduite auprès d'un panel de quatre étudiants volontaires et s'est attachée à dégager des pistes d'amélioration en vue 1) d'optimiser les modalités de formation à la prise en main de l'instrument « carnet » et 2) de collecter de premières traces de situations d'usages et pour éprouver la pertinence des rubriques du carnet mises à disposition des récolteurs et leur potentiel en termes d'auto-récolte. Bien qu'aucun résultat généralisable ne puisse en être déduit, elle offre néanmoins un support de réflexion sur la conduite méthodologique de la phase principale du projet.

5 Pré-analyser les usages au prisme de la modalité d'auto-récolte

La nature des données recueillies dépend des outils qui ont permis la constitution du corpus et nécessite ainsi d'être étudiée au prisme de ceux-ci. Dans le cadre d'une récolte autonome des usages, le mode d'appropriation de l'outil de récolte est documenté afin d'étayer le futur travail d'analyse.

Cette phase de formation puis récolte durant quinze jours réalisée auprès de quatre étudiants volontaires permet par exemple d'établir trois profil-types fondés sur des cri-

tères de situations relatées, de volume du récit ou encore d'éléments de réflexivité présents. Le tableau suivant (cf. Tableau 1, ci-dessous) synthétise ces premières pistes tant au plan de la nature de la récolte que de celui des profils des récolteurs.

Table 1. Croisement des profil-types des étudiants volontaires.

	Profil 1 : « répondre à la demande des chercheuses »	Profil 2 : « explorer les diverses situations d'usage »	Profil 3 : faire le point sur la situation vécue et les conclusions
Temporalité	Carnet rendu au jour 6, deux situations récoltées	4 situations recueillies tout au long de la phase de récolte	Situation unique, remplie au jour 12
Volume du récit	Quelques mots très succincts	Description peu exhaustive mais qui rend intelligible le fond de la situation	Décrit largement la situation, les buts de son activité, les ressources à disposition et les moyens mis en œuvre ainsi que la manière dont elle « débrouille » la situation
Nature des situations	Comparaison entre le résultat attendu et le résultat obtenu (ex. navigation superflue)	Diversité de difficultés liées à l'utilisation des outils (fonctionnalités difficiles à prendre en main) et à leur efficacité par rapport aux résultats attendus	Satisfaction à la première utilisation d'un outil initialement perçu comme complexe
Activité suscitée	Futurs imaginés en termes d'amélioration des outils	Réflexion conduite sur les configurations d'usage (ex. collaboratives, individuelles)	Saisit l'occasion de l'auto-récolte pour faire le point sur les conclusions tirées de la situation
Exemple de verbatim	*« Le résultat me met face à un ouvrage entier alors que je suis seulement à la recherche d'une définition »*	*« Il faudrait un site pour partager et modifier plus facilement notre travail en simultané »*	*« J'ai noté sur mon cahier qu'il valait mieux mettre des mots en anglais pour la récupération d'images pour illustrer le Kahoot »*

Le modeste corpus de données constitué par les quatre volontaires de la phase exploratoire met déjà en évidence une diversité de formes d'appropriation de l'outil de récolte. Des déclencheurs de l'activité de récolte autonome peuvent également être inférés à partir de traces d'un registre affectif qui emprunte différentes formes (profil 1 : « *ça m'énerve que* l'appli soit comme cela » ; profil 2 : « ça s'est *mieux* passé que je l'aurais

pensé, je m'en suis sortie »). Ils apportent de premiers exemples de situations marquées par le rapport à un outil qui ne produit pas les résultats attendus par le sujet en termes de fonctionnement (navigation trop longue, impossible d'être directement sur la messagerie…) ou de résultats (souhait d'obtenir une définition et non un livre).

La récolte est investie dans divers objectifs (exprimer des recommandations pour la conception, faire état des futures modalités d'usage à prévoir ou encore mettre en mots les apprentissages réalisés à l'occasion de la situation vécue) et la mise en récit fait parfois l'objet d'un adressage à autrui, non anticipé initialement.

La démarche autonome de récolte est déjà une première occasion pour le sujet d'expliciter [1] le vécu, voire de faire le point sur la transformation de ses ressources. Hypothèse peut être faite que les situations d'usage récoltées sont significatives pour le sujet car elles l'ont interpellé et potentiellement conduit à sortir du « cours habituel de son activité » [6, 7] voire à reconfigurer son agir.

6 Conclusion : des pistes pour former les récolteurs

L'accès aux usages auto-récoltés est une trappe d'accès aux moyens et conditions par lesquelles le sujet infère des propriétés à un artefact en vue d'agir et d'apprendre [2, 4, 5], utiles à constituer en référence dans les projets de conception [2] et en ce sens indispensables aux chercheurs et développeurs d'applicatifs informatiques.

La primo-analyse présentée dans cette contribution offre des perspectives scientifiques et méthodologiques en termes de supports alternatifs à la vidéo pour la confrontation aux traces d'activité. Une activité médiatisée peut être suscitée à ces fins et doit faire l'objet d'une prise en main centrée sur l'interpellation ou l'étonnement en situation, mais aussi être « reprise » dans un potentiel espace de mise en partage des situations récoltées afin de favoriser une capitalisation des situations vécues et permettre aux sujets de se reconnaître comme concepteurs-dans-l'usage [2].

Références

1. Vermersch, P.: Aide à l'explicitation et retour réflexif. Education permanente 160, 71-80 (2004).
2. Folcher, V.: Développement des hommes et des techniques, perspectives de conception pour et dans l'usage. Document de synthèse en vue de l'Habilitation à Diriger des Recherches. Université Paris 8 (2010).
3. Mayen, P., Métral, J.-F., Tourmen, C.: Les situations de travail. Références pour les référentiels. Recherche et formation 64, 31-46 (2010).
4. Rabardel, P.: Qu'est-ce qu'un instrument ? Appropriation, conceptualisation, mises en situation. CNDP-DIE, pp. 61-65 (1995).
5. Couillaud, S.: Développement du pouvoir d'agir des acteurs en co-innovation par les usages : l'exemple d'accompagnement du projet d'habitat évolutif pour seniors. Thèse de doctorat en ergonomie. Université Paris 8 (2018).
6. Dewey, J.: Logique. La théorie de l'enquête. PUF, Paris (1938).
7. Barbier, J.-M.: Un nouvel enjeu pour la recherche en formation : entrer par l'activité. Savoirs, 33, 9-22 (2013).

L'Identification des Projets de Logiciel Libre Accessibles aux Nouveaux Contributeurs

Paul Hervot[1] et Benoît Crespin[2]

[1] Laboratoire de Recherche de L'EPITA (LRE), 14-16 rue Voltaire, 94270 Le Kremlin-Bicêtre, France
[2] Université de Limoges, XLIM/ASALI, UMR CNRS 7252, France

Résumé Le logiciel libre prend de plus en plus de place dans le paysage public et industriel, notamment pour sa transparence et sa gestion démocratique, pour autant publier le code source d'un logiciel ne le rend pas automatiquement accessible, et les nouveaux contributeurs de ces projets rencontrent de nombreuses barrières les entravant dans leurs contributions. Au travers d'une analyse à grande échelle de l'archive de Software Heritage, nous testons la pertinence de trois indicateurs dans l'identification des logiciels libres accessibles aux nouveaux contributeurs. Nos résultats montrent une corrélation positive entre le nombre de premières contributions réussies dans un projet et la présence d'instructions de contribution, ainsi qu'entre ce même nombre et celui des contributeurs uniques récents du projet. Ces indicateurs trouveront une utilité dans l'enseignement des pratiques du logiciel libre pour aider les enseignants à sélectionner des projets accessibles pour leurs étudiants.

Mots-clé : Collecte, traitement et analyse des traces d'apprentissage · Analyse d'usage(s) et de pratiques · Logiciel libre · Analyse automatique de dépôts logiciels · Barrières d'entrée du logiciel libre

Abstract. FOSS makes an increasing amount of the public and industrial software landscape, notably for its transparency and democratic governance. However, simply publishing the source code of a software does not automatically make it accessible, and many barriers impede new contributors approaching these projects. Through a large-scale software mining of the Software Heritage archive, we test the pertinence of three signals in the identification of accessible FOSS projects for new contributors. Our results show a positive correlation between the number of new contributors of a project successfully bringing their contribution to completion and the presence of contributing guidelines, as well as between that same number and the number of recent unique contributors in the project. Such signals could find a use in the teaching of FOSS practices, helping teachers to select accessible projects for their students.

Keywords: Collection, processing and analysis of learning traces · Usage and practices analysis · FOSS · Mining Software Repositories · Open source barriers to entry

1 Introduction

Nous nous intéressons dans cet article aux indicateurs permettant d'identifier les logiciels libres les plus accessibles aux nouveaux contributeurs, de façon à aider les enseignants à sélectionner des projets sur lesquels faire travailler leurs étudiants. Nous présentons dans la section suivante les travaux existants autour de la notion de nouveaux contributeurs. Après avoir décrit rapidement l'archive de Software Heritage choisie pour notre étude, nous testons la pertinence de trois indicateurs dans l'identification des logiciels libres accessibles.

2 Nouveaux contributeurs

Mendez et al. [4] notent la difficulté des nouveaux volontaires à rejoindre une communauté de logiciel libre, citant comme exemple extrême le projet Apache Hadoop qui a vu 82% de ses nouveaux volontaires quitter le projet après leur première contribution [12]. Plusieurs mesures *a posteriori* de l'accessibilité des projets de logiciel libre ont été faites, en majorité de façon qualitative [11, 9]. Une approche quantitative encore rare consiste à étudier la progression du nombre total de contributeurs au cours de la vie du projet [2]. Similairement, Qiu et al. [6] étudient les signaux que les potentiels nouveaux contributeurs observent pour choisir un projet auquel contribuer. Ils s'intéressent pour cela spécifiquement aux projets disponibles publiquement sur GitHub, ce qui leur permet d'étudier empiriquement l'effet des signaux que la plateforme met en évidence. Ils suggèrent qu'une mesure possible de l'accessibilité pour les nouveaux contributeurs (« *newcomers openness* » dans l'article) est le pourcentage de *pull requests* créées par des contributeurs externes. Pour leur analyse quantitative, Qiu et al. [6] collectent leurs données via l'*API* de GitHub, mais les limites de cette approche font l'objet d'une littérature grandissante et remettant en question certains résultats [1, 13]. Steinmacher et al. [10] remarquent de plus que dans la littérature, les études se concentrent trop sur les projets importants et matures. À l'inverse, 71% des projets hébergés sur GitHub sont personnels et non réellement collaboratifs, Kalliamvakou et al. [1] ne retiennent donc quant à eux que ceux ayant au moins deux auteurs uniques différents.

3 L'archive de logiciels de Software Heritage

Software Heritage est une initiative exploitant des techniques avancées de compression de graphe afin de construire une archive aussi complète que possible du code source actuellement disponible publiquement dans le monde. L'archive est publique et comptait en 2018 plus d'un milliard de *commits* uniques archivés depuis 85 millions d'origines différentes. Elle s'étoffe continuellement à mesure que des origines y sont ajoutées ou revisitées, ce qui en fait l'un des corpus les plus complets et exploitables par la recherche scientifique [5, 8].

4 Méthodologie

Qiu et al. [6] se sont intéressés aux signaux que les potentiels nouveaux contributeurs observent pour choisir un projet auquel *essayer* de contribuer, nous proposons de déterminer si certains de ces signaux sont de surcroit prédictifs d'une réelle accessibilité de ces projets, c'est à dire à quel point de nouveaux contributeurs *réussissent* à produire une contribution apparaissant dans l'historique de développement du projet. Pour dépasser les limitations de GitHub, nous utiliserons l'archive de Software Heritage comme accès à la population étudiée. Ce choix nous empêchera cependant d'étudier les indicateurs propres à GitHub comme le nombre de *pull requests* d'un projet. Pour évaluer l'accessibilité d'un projet, nous proposons d'utiliser comme variable proxy le nombre de contributeurs apparaissant pour la première fois dans l'historique de développement du projet entre le premier juin 2019 et le premier septembre 2019 [6]. Nous considérons comme « récent » tout événement survenu dans les six mois avant la période de référence étudiée.

Hypothèse 01 *: les projets possédant des instructions de contribution sont plus accessibles pour les nouveaux contributeurs que ceux n'en ayant pas.*

Hypothèse 02 *: le nombre de contributeurs uniques récents d'un projet est positivement corrélé à son accessibilité pour les nouveaux contributeurs.*

Hypothèse 03 *: le nombre de* commits *récents au sein d'un projet est positivement corrélé à son accessibilité pour les nouveaux contributeurs.*

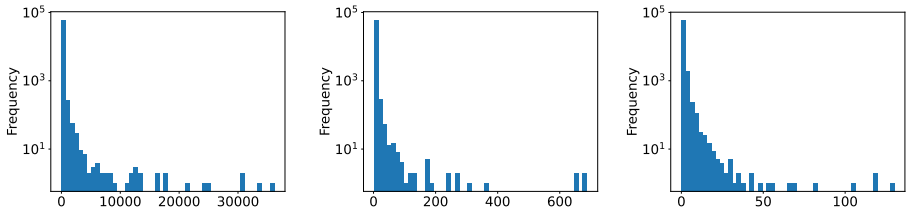
Sont exclus de l'échantillon les projets : 1. qui sont des *forks* d'un autre (en cas de *commits* en commun, seul le projet qui la plus longue chaîne de *commits* a été retenu comme représentant du groupe) ; 2. n'ayant reçu aucun *commit* pendant la période de référence (inactivité) [1] et 3. ayant moins de deux contributeurs uniques récents (projet personnel non-collaboratif) [1].

Nous effectuons un premier parcours du graphe à partir de chaque origine afin d'identifier le représentant de chaque groupe de fork. Pour le *snapshot* le plus récent de chaque origine ainsi retenue, un nouveau parcours est effectué à partir de sa branche principale afin de récolter les données de recherche. La présence d'instructions de contribution en particulier est difficile à vérifier car le graphe ne contient que les noms et la hiérarchie des fichiers d'un projet, pas leur contenu. Si un fichier `CONTRIBUTING.md` ou assimilé existe, nous validons la présence d'instructions de contribution, sinon, nous cherchons un fichier `README.md` ou assimilé. En l'absence d'un tel fichier, nous concluons à l'*absence* d'instructions de contribution, en sa présence, nous sauvegardons l'identifiant unique du fichier afin d'analyser son contenu ultérieurement. Celui-ci peut être téléchargé publiquement en HTTP sur le site archive.softwareheritage.org ou depuis le *registry* Amazon S3 software-heritage. Le site limite le débit des requêtes autorisées, contrairement au *registry*, mais ce dernier est incomplet. Nous téléchargeons donc en priorité via le *registry* Amazon S3 et nous nous rabattons sur le site pour les fichiers manquants. Nous cherchons ensuite dans le `README` une section contenant le mot *contributing* ou une expression proche. Les noms de sections sont identifiés

en supposant que les `README` sont formatés en Markdown ou reStructuredText [3, 7]. Enfin, le nombre de nouveaux contributeurs est calculé en comptant les contributeurs uniques de la période de référence qui n'apparaissent dans aucun *commit* antérieur à cette période, et le nombre de contributeurs uniques récents en les comptant simplement dans la période de six mois précédent la période de référence.

Un *replication package* contenant le code source et les bibliothèques utilisés pour l'étude présentée dans notre article, ainsi que les données brutes collectées aux différentes étapes, est disponible sur Zenodo : zenodo.org/record/7888415.

5 Résultats et discussion

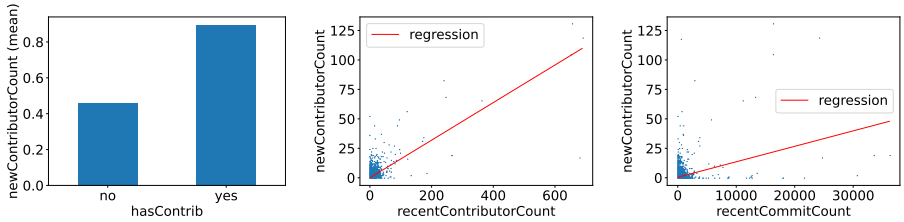


(a) *Commits* récents (b) Contributeurs récents (c) Nouveaux contributeurs

Fig. 1 – Distributions (avec ordonnée logarithmique) des données étudiées

Sur les 60966 projets distincts (sans historique commun) analysés, 14% d'entre eux possèdent des instructions de contribution. Les nombres de *commits* récents, de contributeurs récents et de nouveaux contributeurs varient fortement au sein de cette population, la distribution n'est normale dans aucun des cas (voir Fig. 1). Les projets possédant des instructions de contribution ont en moyenne plus de nouveaux contributeurs (voir Fig. 2a) avec une taille d'effet $\rho \approx 0.57$ (test MWW), ce qui confirme ($p \approx 0$) que les deux distributions sont bien différentes et valide l'hypothèse H01. De futures recherches pourraient essayer de déterminer si ce plus grand nombre de nouveaux contributeurs *ayant réussi* à contribuer est dû uniquement à ce plus grand nombre de nouveaux contributeurs *essayant* de contribuer, ou si la présence d'instructions de contribution a un réel rôle dans le succès d'une tentative de contribution. Notons tout de même que la robustesse du test MWW diminue significativement dans une situation comme la nôtre (distribution non-normale et variances significativement différentes entre les deux catégories) [14].

Une régression GLS suggère que plus le nombre de contributeurs récents d'un projet est élevé, plus son nombre de nouveaux contributeurs l'est aussi (voir Fig. 2b). Le coefficient de détermination $R^2 \approx 0.45$ du modèle indique que le nombre de contributeurs récents explique environ 45% de la variation du nombre



(a) Présence d'instructions de contribution

(b) Nombre de contributeurs récents

(c) Nombre de *commits* récents

Fig. 2 – Nombre de nouveaux contributeurs en fonction des variables mesurées.

de nouveaux contributeurs, ce qui valide l'hypothèse H02, avec une taille d'effet tout de même modérée. Suivant la même approche pour l'analyse du nombre de *commits* récents, une régression GLS suggère ici aussi une corrélation positive entre le nombre de *commits* récents d'un projet et son nombre de nouveaux contributeurs (voir Fig. 2c), mais son coefficient de détermination est trop faible ($R^2 \approx 0.10$) pour considérer que le modèle représente fidèlement des données du problème. Nous ne pouvons donc valider l'hypothèse H03.

6 Conclusion

L'analyse de l'archive de Software Heritage nous a permis de trouver une corrélation positive entre le nombre de nouveaux contributeurs d'un projet de logiciel libre et deux indicateurs facilement observables pour les aspirants contributeurs ou les enseignants souhaitant faire travailler leurs étudiants sur ces problématiques : la présence d'instructions de contribution au sein d'un projet et le nombre de contributeurs uniques ayant récemment contribué au projet. Nous n'avons en revanche pas trouvé de lien avec le nombre de *commits* récents des projets. Restent par ailleurs à identifier les liens de causalité expliquant nos observations, ainsi que la mise en situation de ces indicateurs dans un contexte pédagogique.

Références

1. E. Kalliamvakou et al. The Promises and Perils of Mining GitHub. In : *Proceedings of the 11th Working Conference on Mining Software Repositories*. MSR 2014. Hyderabad, India : Association for Computing Machinery, 2014, p. 92-101. DOI : 10.1145/2597073.2597074.
2. J. Lerner, P. A. Pathak et J. Tirole. The Dynamics of Open-Source Contributors. In : *American Economic Review* 96.2 (mai 2006), p. 114-118. DOI : 10.1257/000282806777211874.
3. Markdown Basic Syntax - Headings. URL : www.markdownguide.org/basic-syntax/#headings.

4. C. Mendez et al. Open Source Barriers to Entry, Revisited : A Sociotechnical Perspective. In : *Proceedings of the 40th International Conference on Software Engineering*. ICSE '18. Gothenburg, Sweden : Association for Computing Machinery, 2018, p. 1004-1015. DOI : 10.1145/3180155.3180241.
5. A. Pietri, D. Spinellis et S. Zacchiroli. The Software Heritage Graph Dataset : Public Software Development Under One Roof. In : *2019 IEEE/ACM 16th International Conference on Mining Software Repositories (MSR)*. 2019, p. 138-142. DOI : 10.1109/MSR.2019.00030.
6. H. S. Qiu et al. The Signals That Potential Contributors Look for When Choosing Open-Source Projects. In : *Proc. ACM Hum.-Comput. Interact.* 3.CSCW (nov. 2019). DOI : 10.1145/3359224.
7. reStructuredText Syntax - Sections. URL : docutils.sourceforge.io/docs/ref/rst/restructuredtext.html#sections.
8. G. Rousseau, R. Di Cosmo et S. Zacchiroli. Growth and Duplication of Public Source Code over Time : Provenance Tracking at Scale. working paper or preprint. Juin 2019. URL : hal.archives-ouvertes.fr/hal-02158292.
9. C. Stanik et al. A Simple NLP-Based Approach to Support Onboarding and Retention in Open Source Communities. In : *2018 IEEE International Conference on Software Maintenance and Evolution (ICSME)*. 2018, p. 172-182. DOI : 10.1109/ICSME.2018.00027.
10. I. Steinmacher et al. A systematic literature review on the barriers faced by newcomers to open source software projects. In : *Information and Software Technology* 59 (2015), p. 67-85. DOI : 10.1016/j.infsof.2014.11.001.
11. I. Steinmacher et al. Overcoming Open Source Project Entry Barriers with a Portal for Newcomers. In : *Proceedings of the 38th International Conference on Software Engineering*. ICSE '16. Austin, Texas : Association for Computing Machinery, 2016, p. 273-284. DOI : 10.1145/2884781.2884806.
12. I. Steinmacher et al. Why do newcomers abandon open source software projects ? In : *2013 6th International Workshop on Cooperative and Human Aspects of Software Engineering (CHASE)*. 2013, p. 25-32. DOI : 10.1109/CHASE.2013.6614728.
13. M. Z. Trujillo, L. Hébert-Dufresne et J. Bagrow. The penumbra of open source : projects outside of centralized platforms are longer maintained, more academic and more collaborative. In : *EPJ Data Science* 11.1 (2022), p. 31.
14. D. W. Zimmerman. Invalidation of Parametric and Nonparametric Statistical Tests by Concurrent Violation of Two Assumptions. In : *The Journal of Experimental Education* 67.1 (1998), p. 55-68. DOI : 10.1080/00220979809598344.

EIAH2023 : 11ème Conférence sur les Environnements Informatiques pour l'Apprentissage Humain
La conférence pluridisciplinaire francophone sur la conception et l'analyse des environnements numériques pour l'éducation et la formation
12-16 juin 2023 Brest (France)

Session de communications 4.2
Retours d'usage et évaluation

Comprendre les usages et effets d'un système de recommandations pédagogiques en contexte d'apprentissage non-formel

Anaëlle Badier, Mathieu Lefort, Marie Lefevre

Univ Lyon, UCBL, CNRS, INSA Lyon, LIRIS, UMR5205,
F-69622 Villeurbanne, France
prenom.nom@liris.cnrs.fr

Résumé. Nos travaux se situent dans un contexte d'apprentissage non-formel sur application mobile où nous avons proposé un moteur de recommandation de ressources pédagogiques. Les apprenants concernés ont des objectifs et modalités de travail très variés. Dans cet article, après avoir rappelé le fonctionnement de notre moteur de recommandations, nous analysons les traces d'utilisation de plus de 10 000 apprenants sur 4 mois, montrons que les recommandations répondent à des objectifs et des usages différents, et que leur suivi influe positivement sur l'expérience d'apprentissage.

Mots-clés : Recommandations pédagogiques , Analyse de Traces , Analyse d'usages et de pratiques , Apprentissage auto-régulé

Abstract. Our work takes place in a non-formal learning context on a mobile application in which we propose an educational recommendation system. Our learners have very different objectives and learning behaviours. In this article, after having briefly explained the recommendation system, we analyze the learning traces of more than 10,000 learners over 4 months, show that the recommendations meet different objectives and uses, and that following them has a positive impact on the learning experience.

Keywords: Pedagogical Recommendations , Traces Analysis , Usage and Practice Analysis , Self-Regulated Learning

Le nombre croissant des applications et plateformes de e-learning depuis plusieurs décennies a conduit à de nouvelles formes d'apprentissage. Les apprenants sont de plus en plus amenés à travailler par eux-mêmes en recherchant des contenus d'apprentissage hors de la classe. Dans ce contexte, nous travaillons avec une application mobile parascolaire qui propose des contenus de révision pour tous les niveaux scolaires et toutes les disciplines, y compris les enseignements professionnels ou de spécialités. Au sein de cette application, la navigation est libre et l'utilisation n'est pas encadrée. De ce fait, les apprenants l'utilisent comme

un outil de micro-apprentissage, avec un travail occasionnel, non régulier et des sessions de travail très courtes.

Nous pensons que la mise en place d'un système de navigation guidée personnalisée au sein de l'application peut améliorer l'expérience utilisateur. Pour ce faire, nous avons implémenté un système de recommandations pour les Mathématiques du Collège au Lycée, pédagogiquement validé par des experts et utilisé par les apprenants [1]. Dans cet article, nous nous intéressons à l'expérience utilisateur avec ce système au travers des questions de recherche suivantes :

Q1- Quels facteurs influencent le suivi des recommandations par les apprenants dans un contexte parascolaire ?

Q2 - Quels usages des recommandations pouvons-nous observer dans une application parascolaire ?

Après avoir brièvement rappelé le fonctionnement du système de recommandation présenté dans l'application et l'avoir positionné au regard des travaux du domaine (section 1), nous présentons le cadre d'expérimentation et de recueil des données d'utilisation (section 2). Dans cette étude exploratoire, nous analysons ensuite selon plusieurs critères les recommandations suivies par les apprenants parmi celles proposées (section 3) et finissons notre étude sur les modifications observées dans l'expérience d'apprentissage (section 4). Les résultats présentés sont discutés dans la section 5.

1 Recommander en contexte d'apprentissage non-formel

L'application mobile avec laquelle nous travaillons organise les ressources pédagogiques au sein de disciplines (Mathématiques, Français, Anglais...), elles-même subdivisées en chapitres (Arithmétique, Droites du plan, Fonctions de références...). Au sein de chaque chapitre, l'élève a à disposition des mini-cours (1 à 5), des fiches de révision et 4 quiz de 5 questions au format QCM. Nous proposons un moteur de recommandations pour accompagner l'élève de chapitres en chapitres. Considérant un étudiant qui travaille sur un chapitre donné (*Chapitre d'entrée*), nous recommandons une petite liste de chapitres (*Chapitres recommandés*) parmi les chapitres disponibles dans l'application. Selon la classification de Brusilovsky [5], il s'agit d'un système de support de navigation adaptatif. Le fonctionnement général du système est présenté sur la figure 1. Pour proposer des recommandations et accompagner l'apprenant dans son utilisation de l'application, le système est basé sur 3 modules : un **Module Pédagogique**, un **Module Historique** et un **Module Nouveauté**.

1.1 Sélection et classement pédagogiques des ressources

Le **Module Pédagogique** décrit dans [1] sélectionne une stratégie de recommandation (3 sur la figure 1), filtre les *Chapitres Candidats* selon cette stratégie d'apprentissage (4) et calcule un score de pertinence pédagogique pour classer les chapitres et recommander les meilleurs (5).

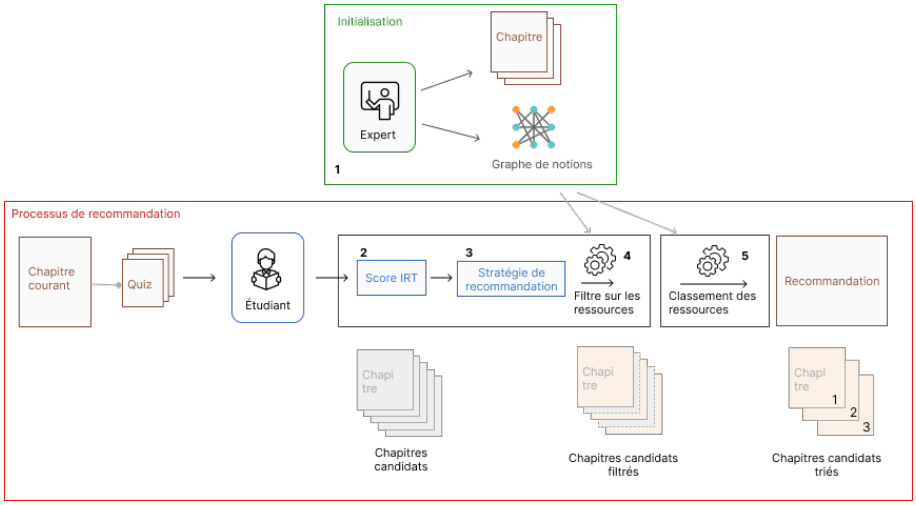


Fig. 1. Fonctionnement global du système de recommandations. Les étapes numérotées sont référencées et expliquées ci-dessous.

Les chapitres disponibles dans l'application sont organisés dans un graphe de notions (1). D'autres travaux organisent leurs ressources avec des approches axées compétences comme dans le projet ComPer [16] ou utilisent des ontologies [9]. Dans notre cas, nous ne recommandons que les ressources internes à l'application, qui sont étiquetées avec des notions clés par les professeurs concepteurs de ces ressources. Nous appelons *notion* une composante de savoir abordée dans un chapitre. Une notion est étiquetée comme *prérequise* au chapitre si l'apprenant doit avoir déjà rencontré cette notion précédemment pour aborder le chapitre en question. Une notion est attendue si le chapitre permet de découvrir cette notion ou d'aller plus loin dans sa maîtrise. Avec ce système de notions clés, les chapitres sont liés les uns aux autres. Un chapitre est d'autant plus proche d'un autre qu'il partage des notions spécifiques en commun.

Nous définissons 3 stratégies de recommandations : révision, continuité et approfondissement. Ces stratégies sont assignées à l'apprenant à la fin de l'étude de chaque chapitre selon sa maîtrise du chapitre terminé. Plusieurs systèmes de recommandation se basent sur le concept de Zone Proximale de Développement (ZPD) [20] pour attribuer des ressources à des apprenants basées sur le niveau de difficulté de la ressource et le niveau de compétence estimé de l'apprenant [3]. D'autres approches utilisent les Modèles de Traçage de Connaissances *(Knowledge Tracing Models, KTM)* pour modéliser le niveau de connaissance d'un apprenant et lui recommander du contenu adapté. Parmis ces *KTM*, on peut citer le *Bayesian Knowledge Tracing, BKT* de Corbett et Anderson [8] ou d'autres systèmes *KTM* comme ceux de Vie et Kashima pour prédire la performance des apprenants dans le cadre des MOOCs [18]. Dans notre cas, nous estimons le niveau de maîtrise d'un chapitre à partir des quiz effectués par l'élève

dans ce chapitre. En utilisant le modèle de la théorie de réponse à un item (*Item Response Theory, IRT*) développé par Baker [2], nous calculons un score θ qui représente le niveau de maîtrise des quiz, et par extension du chapitre (2). Nous définissons 3 groupes de niveaux selon le score θ et attribuons les stratégies de recommandation selon ces groupes (3). Les stratégies étant ainsi définies, et recalculées pour chaque élève à chaque nouveau chapitre, nous filtrons les chapitres éligibles à la recommandation (4). Nous calculons ensuite un score de pertinence pédagogique (cf. équation 1) pour chaque chapitre éligible, basé sur les notions en commun ($score_{similarite}$) et la distance ($penalite_{distance}$) au niveau scolaire du chapitre courant, ce qui permet de classer les chapitres éligibles (5). Les détails du filtrage et du calcul du score pédagogique sont présentés dans l'article [1]. Les recommandations fournies par ce système ont été validées par les experts pédagogiques et sont utilisées par les apprenants.

$$score_{pedagogique} = score_{similarite} * (1 - penalite_{distance}) \tag{1}$$

Prendre en compte l'historique de l'apprenant. Le système de recommandation étant limité à l'application parascolaire, les seules données *a priori* sur les connaissances des étudiants sont les items explorés au sein de l'application, particulièrement les quiz faits et les résultats obtenus. Nous n'avons aucune information sur les chapitres étudiés en classe. Pour maintenir une variété dans l'activité au sein de l'application, nous ne souhaitons pas recommander des chapitres déjà récemment étudiés et maîtrisés par l'apprenant. Pour cette raison, le **Module Historique** est basé sur deux composantes : la complétion des chapitres candidats et la note obtenue sur ces chapitres. Le module historique évolue avec le temps. En effet, la dimension temporelle est un facteur important dans de nombreux systèmes de recommandation : Choffin *et al.* ont étudié l'effet de la répétition sur l'apprentissage et l'oubli [6], et Jiang *et al.* utilisent un modèle de réseau de neurones LSTM pour inclure un paramètre d'oubli [12]. La distribution de la rétention mémorielle au cours du temps, décrite par Ebbinghaus [10] a été utilisée dans de nombreux travaux depuis [19]. Nous introduisons dans notre module historique un facteur de rétention w_t, calculé à partir de l'allure de la courbe d'oubli d'Ebbinghaus, mais adaptée pour notre contexte d'usage non régulier : dans notre cas, la courbe d'oubli estime une rétention de 70% après 7 jours, diminuée à 12% après 30 jours. Le score du module historique est calculé avec l'équation 2.

$$score_{historique} = 1 - [w_t * score_{completion} * score_{note}] \tag{2}$$

avec $w_t \in [0, 1]$ le facteur de rétention, $score_{completion} \in [0, 1]$ le ratio de quiz terminés dans le chapitre candidat et $score_{note} \in [0, 1]$ la note moyenne sur les quiz terminés dans le chapitre candidat.

Avec ce module, nous prenons en compte les ressources déjà étudiées sur l'application. Nous nous autorisons toutefois à recommander du contenu non terminé ou non maîtrisé, ou étudié longtemps avant la recommandation, en supposant qu'un rappel de ces chapitres peut être bénéfique. Ce module intervient dans la partie classement des ressources (5 sur la figure 1).

Encourager la curiosité en variant les recommandations. Notre système de recommandation est appliqué à un contexte d'apprentissage non formel. Pour permettre un renouvellement de l'offre de recommandations et palier un effet de lassitude, nous introduisons un **Module Nouveauté**. Oudeyer *et.al* ont démontré que la curiosité et la nouveauté sont importants pour améliorer l'apprentissage et la rétention de l'information [14]. La sérendipité est également un critère notable pour rendre les recommandations attractives [23]. Comme pour le module historique, nous conservons un facteur temporel pour nous permettre de recommander des ressources déjà proposées longtemps auparavant. Pour limiter le nombre de paramètres dans le modèle dans un premier temps, et sans études approfondies sur l'évolution de la rétention mémorielle dans un contexte de navigation sur application mobile, nous posons que le facteur temporel w_t est identique à celui utilisé dans le module historique. Le score de nouveauté est calculé avec l'équation 3. Ainsi, une ressource déjà fortement recommandée se retrouve pénalisée dans son score de nouveauté. Ce module intervient dans la partie classement des ressources (5 sur la figure 1).

$$score_{nouveaute} = 1 - [w_t * \frac{1}{n} \sum_{i=1}^{n} \frac{3 - k_i}{3}] \tag{3}$$

avec n le nombre de fois où le chapitre a été recommandé, t la dernière fois que le chapitre a été recommandé et $k_i \in [0, 3[$ le rang de la ressource dans le top-3 des recommandations proposées (0 étant le meilleur) la i^e fois.

Pour paramétrer notre modèle de recommandations et faire le classement (5), nous décidons de donner plus d'importance au module pédagogique, puis au module historique et enfin au module nouveauté. Cette initialisation manuelle des poids initiaux de l'équation 4, semblable a ce qui a pu être fait par Zhao *et. al* [22], est discuté dans la section 5.

$$score_{reco} = 0.5 * score_{pedagogique} + 0.4 * score_{historique} + 0.1 * score_{nouveaute} \tag{4}$$

2 Protocole d'évaluation

De nombreux systèmes de recommandation sont utilisés et évalués en classe [15,17,21], ou avec des étudiants simulés virtuellement [7] pour disposer d'une quantité de données suffisante pour entraîner les modèles d'apprentissage, ce qui n'est pas notre cas. En raison de notre contexte et pour comprendre les besoins et comportements de nos apprenants, nous évaluons directement le système dans les conditions réelles d'utilisation. Le principal inconvénient de ce système de validation "en vie réelle" est la difficulté de collecte de données comme souligné par Erdt *et. al* [11]. Cette difficulté est accentuée dans les contextes d'apprentissage auto-régulés, puisque les apprenants peuvent quitter à tout moment la plateforme d'apprentissage.

Le moteur de recommandation a été implémenté pour la discipline Mathématiques, pour les niveaux scolaires de la 6^e à la Terminale générale. Il a été mis à la disposition de tous les nouveaux apprenants pour une durée de 1 mois à

partir de leur date d'inscription sur l'application. Les recommandations ont été proposées *via* 3 points d'entrée :

— *Quiz échoué.* Si l'apprenant échoue totalement à l'un des quiz de 5 questions du chapitre, un bouton lui propose d'accéder directement à des recommandations de révision.
— *Parcours interrompu.* Les apprenants sont incités à travailler avec 4 quiz de 5 questions. Si un apprenant interrompt son parcours, une fenêtre propose de continuer à travailler avec des recommandations dont la stratégie a été attribuée selon les réponses déjà données.
— *Fin du parcours.* Les recommandations sont affichées sur la page de résultat à la fin du parcours de 4 quiz et dépendent des notes obtenues sur ces 4 quiz.

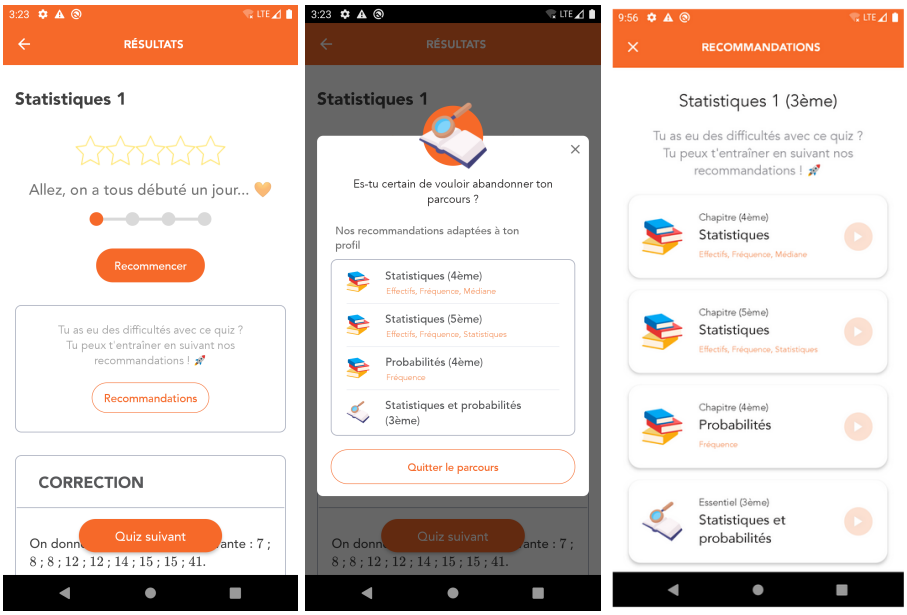


Fig. 2. Interface de recommandations pour le chapitre *Statistiques* de niveau 3ème pour la stratégie révision. À gauche, par le point d'entrée quiz échoué, au milieu le cas d'un parcours interrompu et à droite l'interface de fin du parcours. Le niveau scolaire des chapitres recommandés est affiché entre parenthèses. Les notions en commun avec le chapitre actuel sont indiquées en orange. Le contenu "Essentiel" ne fait pas partie du système de recommandation présenté dans le cadre de cet article.

L'interface de présentation des recommandations est présenté sur la figure 2. Pour analyser l'utilisation faite par les apprenants des recommandations proposées, nous mettons en place un mécanisme de collecte des traces d'apprentissage spécifique. Semblable au système xAPI utilisé dans plusieurs travaux [13,4], chaque action de l'utilisateur est identifiée et stockée en base. Nous pou-

vons ainsi identifier les actions de *Vue de l'interface de recommandations*, *Clic sur une recommandation de l'interface*, *Quitter l'interface de recommandations* mais également des événements plus génériques comme *Connexion à l'application*, *Quiz commencé*, *Mini-cours consulté*, ... avec les informations de contexte associées : niveau scolaire de l'apprenant, discipline de la ressource, identifiant de la session de travail correspondante, heure de la consulation de la ressource, etc. Les apprenants sont anonymisés et leur identité n'est pas stockée dans les bases de données d'analyse de traces pour des raisons de protection des données. Nous présentons dans les sections suivantes les résultats observés pendant la période du 07 septembre 2022 au 07 janvier 2023.

3 Choix des recommandations par les apprenants

Nous avons collecté les traces d'utilisations de 11 962 apprenants ayant été confrontés au moins une fois à une recommandation, ce qui représente 39 420 traces de recommandation vues, dont 5 609 recommandations suivies par 3 448 apprenants. Pour répondre à notre première question de recherche sur les facteurs influençant le suivi des recommandations, nous avons analysé les données selon plusieurs critères.

Table 1. Influence de la source de la recommandation sur la décision de suivi (test du Chi-2 de Pearson. $\chi^2 = 2714.6$, df=2 , $V = 0.3$, $p_{value} < 2.2e-16$).

Point d'entrée	Recommandation *effectifs observés (résidus)*	
	Choisie	Rejetée
Quiz échoué	904 (+34.9)	1202 (-14.2)
Parcours interrompu	1724 (-26.6)	21041 (+10.8)
Fin du parcours	2981 (+20.0)	11568 (-8.2)

Le premier critère étudié est le point d'entrée de l'affichage des recommandations. Les résultats présentés sur le tableau 1 montrent que les recommandations sont plus suivies dans un contexte de quiz échoué (résidus +34.9) ou de parcours terminé (résidus +20). Les recommandations issues de parcours interrompus sont plus rejetées que les autres (résidus -26.6).

Le second critère est l'influence de la stratégie de recommandation appliquée sur le choix de suivre ou non une recommandation. Les résultats sont présentés tableau 2. Nous constatons que les recommandations associées à la stratégie de révision ont tendance à être plus suivies (résidus +10.9) que les stratégies de continuité ou d'approfondissement.

Ces premières analyses montrent que le suivi des recommandations dépend du moment où elles sont proposées (quiz échoué, fin du parcours) et qu'elles sont davantage adoptées lorsqu'elles proposent des contenus de révision.

Table 2. Influence de la stratégie de recommandation sur la décision de suivi (test du Chi-2 de Pearson, $\chi^2 = 207.19$, df=2 , $V = 0.1$, $p_{value} < 2.2e-16$).

Stratégie	Recommandation *effectifs observés (résidus)*	
	Choisie	Rejetée
Révision	1682 (+10.9)	7384 (-4.4)
Continuité	1815 (-7.6)	13431 (+3.1)
Approfondissement	2112 (-0.8)	12996 (+0.3)

4 Effets des recommandations sur l'expérience d'apprentissage

Après avoir identifié certains critères de sélection des recommandations, et afin de traiter notre deuxième question de recherche adressant les usages observés des recommandations, nous étudions les activités réalisées au sein des chapitres recommandés suivis. Pour cela, nous reprenons les 5 609 traces des 3 448 étudiants ayant suivi au moins une fois une recommandation et observons les activités effectuées au sein du chapitre qu'ils ont choisi. Les résultats sont présentés dans le tableau 3.

Table 3. Activités réalisées au sein des chapitres pour les ressources recommandées et suivies. Une recommandation suivie peut à la fois mener à une activité de quiz commencé, quiz terminé et cours consulté (total $> 100\%$).

Activité dans le chapitre suivi	Nombre de recommandations concernées N=5609
Aucune activité	1813 (32.3%)
Au moins 1 cours consulté	1668 (29.7%)
Au moins 1 quiz commencé	3061 (54.6%)
Au moins 1 quiz commencé et terminé	2027 (36.1%)

Au sein de chaque chapitre, l'apprenant peut étudier 2 ressources différentes : les quiz et les cours. Nous constatons que 32.3% des recommandations suivies n'ont débouché sur aucune activité dans le chapitre suivi : soit parce que la recommandation n'a pas correspondu aux attentes des apprenants, ou parce qu'ils ont suivi les recommandations uniquement par curiosité. On pourrait qualifier les recommandations suivies sans activité consécutive dans le chapitre consulté de *suivi consultatif*. Dans la majorité des cas, la recommandation suivie a mené à une réelle activité : les apprenants ont commencé au moins 1 un quiz ou consulté au moins 1 cours (*suivi partiel*) et dans 29.7% ils ont terminé au moins un quiz du chapitre suivi (*suivi approfondi*).

Au-delà de l'activité au sein d'une recommandation suivie, nous nous interrogeons sur l'impact du suivi des recommandations dans l'expérience d'apprentissage. Dans notre contexte d'utilisation volontaire et parascolaire, nous ne cherchons pas à obtenir des gains d'apprentissage mais plutôt des évolutions de comportements de travail avec l'application. Pour qualifier l'expérience d'apprentissage, nous utilisons les indicateurs de *nombre de sessions de travail*, *durée des sessions de travail*, *nombre de quiz commencés*, *nombre de quiz terminés*, *nombre de cours consultés*, *temps moyen de lecture des cours*. Une *session de travail* est définie comme l'ensemble des activités enregistrées dans l'application entre l'ouverture et la fermeture de cette dernière. Un apprenant peut avoir plusieurs sessions de travail dans une même journée.

Nous avons récolté les données de sessions complètes de 1126 apprenants ayant suivi au moins une fois des recommandations, et ayant commencé au moins un quiz de mathématiques, pour comparer l'expérience d'apprentissage entre des sessions avec suivi et sans suivi des recommandations. À l'aide d'un test de Student apparié sur chacun des indicateurs, nous étudions l'impact du suivi des recommandations sur l'expérience d'apprentissage au sein de l'application. Les résultats sont présentés sur le tableau 4.

Table 4. Comparaison par test de Student des indicateurs d'apprentissage entre des sessions avec ou sans suivi des recommandations. ($^{***}pvalue < 0.001, ^{*}pvalue < 0.05$)

Indicateurs par membre	Sessions avec suivi N=1471	Sessions sans suivi N=6222	pvalue
Nombre de sessions	1.31	5.53 ***	<2.2e-16
Temps moyen de session (min)	11.06 ***	6.72	<2.2e-16
Nombre moyen de quiz commencés par session	3.97 ***	2.28	<2.2e-16
Nombre moyen de quiz terminés par session	3.27 ***	1.85	<2.2e-16
Nombre moyen de cours consultés par session	3.01 ***	1.67	<2.2e-16
Temps moyen de lecture des cours par session (sec)	49.4 *	37.9	0.02

Nous constatons que les sessions avec suivi d'une ou plusieurs recommandations sont des sessions de travail plus actives. Les apprenants travaillent presque deux fois plus longtemps, même si nous restons dans un contexte de microlearning. Ils font également plus de quiz, et lisent plus de cours. La conclusion de cette analyse peut être de deux natures : les recommandations encouragent l'apprenant à travailler plus, ou bien lorsque l'apprenant décide de consacrer un temps d'étude plus approfondi avec l'application, il aura tendance à suivre davantage les recommandations proposées. Dans les deux cas, notre système a

un impact positif en proposant une nouvelle expérience d'apprentissage dans l'application.

5 Discussion

Construction du système. Nous avons fait le choix de paramétrer le système en construisant le score de recommandation (équation 4) avec de premières valeurs arbitraires en priorisant le score pédagogique (0.5), puis historique (0.4) et d'accorder peu de poids au score de nouveauté (0.1). Des *ablation studies* auraient été nécessaires pour évaluer précisément l'impact de chacun des modules dans le classement des ressources et le choix des apprenants. De même, le choix de la modélisation du temps avec le paramètre w_t peut être discuté. Des études plus poussées pourraient faire varier ce paramètre pour étudier sa pertinence dans un contexte d'apprentissage sur application mobile. Dans un contexte d'apprentissage parascolaire avec très peu de régularité d'utilisation, nous avons choisi de mener en premier lieu une démarche exploratoire afin d'étudier l'adéquation d'un tel système à notre contexte. Le système a donc été préalablement initialisé pour collecter rapidement de premières données d'utilisation et comprendre les comportements des apprenants.

Influence du suivi sur l'expérience d'apprentissage. Dans un contexte d'utilisation parascolaire et non formel du système de recommandation, il est difficile de comprendre les motivations qui poussent les apprenants à utiliser l'application et nos recommandations d'une certaine manière. Les recommandations semblent être adoptées par les apprenants dans certains contextes (quiz échoué, stratégie de révision) mais moins dans d'autres (parcours interrompu, stratégie d'approfondissement). Quand on s'intéresse à l'influence du suivi des recommandations sur l'expérience d'apprentissage, le tableau 4 montre que le nombre de cours consultés est supérieur pour les sessions avec suivi de recommandations, mais que le temps de lecture n'est que très légèrement supérieur. Nous pouvons poser l'hypothèse que l'apprenant survole les recommandations en passant rapidement sur le contenu recommandé mais sans s'y attarder. En effet, les temps de sessions très courts témoignent bien du caractère très accessoire de l'application qui est utilisée comme un outil support et non comme la première source d'apprentissage. Ce comportement "exploratoire" dans le suivi des recommandations est illustré par le tableau 3 avec des recommandations suivies mais dont les activités pédagogiques relatives ne sont pas systématiquement exploitées. Nous avons choisi d'analyser le comportement des apprenants par le prisme de certains critères et indicateurs, de par le contexte totalement libre de cette forme d'apprentissage, plusieurs autres critères auraient pu être pertinents pour poursuivre l'analyse.

6 Conclusion et Perspectives

Dans cet article nous avons étudié la mise en place d'un moteur de recommandation adapté à un contexte d'apprentissage non formel sur application mobile.

Les recommandations, construites à partir de trois modules (Pédagogique, Historique et Nouveauté), sont présentées aux apprenants à différents moments de leur parcours de travail, et s'articulent autour de 3 stratégies : révision, continuité et approfondissement. Nous avons analysé les traces d'apprentissage et montré que les apprenants utilisent principalement les recommandations du système dans la stratégie de révision, prioritairement lorsqu'ils ont échoué à un quiz ou terminé l'étude du chapitre en cours. Nous avons également mis en évidence différentes modalités de suivi des recommandations (consultatif, partiel ou approfondi) et avons montré que les sessions de travail avec suivi des recommandations étaient plus riches que celles sans suivi. En effet, le temps de session ainsi que le nombre de quiz commencés, terminés et la consultation des cours sont significativement supérieurs dans les sessions avec suivi que dans celles sans suivi des recommandations.

Cette première étude donnant des résultats prometteurs, nous allons poursuivre les analyses du moteur de recommandation appliqué à d'autres disciplines, pour lequel le système est déjà implémenté mais pas encore rendu accessible aux apprenants.

Notre première question de recherche s'intéressait à la sélection des recommandations proposées par les apprenants. Nous allons approfondir l'analyse avec d'autres indicateurs pouvant influencer cette sélection, par exemple l'affichage ou non du niveau scolaire dans les recommandations. En lien avec la première étude publiée dans [1], nous allons examiner si ces critères de sélection sont communs à tous les apprenants, ou varient en fonction des publics.

Notre seconde question de recherche posait la question de l'influence du suivi des recommandations sur l'expérience d'apprentissage. Ayant observé différents niveaux de suivi, dans la suite de nos travaux, nous chercherons à comprendre le comportement de deux publics cibles : ceux qui ne consultent jamais les recommandations proposées et ceux qui ne restent qu'en suivi consultatif. Cette meilleure compréhension permettrait peut-êre de leur proposer des recommandations plus pertinentes.

Références

1. Badier, A., Lefort, M., Lefevre, M. : Recommendation model for an after-school e-learning mobile application. In : Proceedings of CSEDU 15th (2023)
2. Baker, F.B. : The Basics of Item Response Theory. ERIC Clearinghouse on Assessment and Evaluation (2001)
3. Baker, R., Ma, W., Zhao, Y., Wang, S., Ma, Z. : The results of implementing zone of proximal development on learning outcomes. In : The 13th International Conference on Educational Data Mining (2020)
4. Branthôme, M. : Pyrates : A Serious Game Designed to Support the Transition from Block-Based to Text-Based Programming. In : EC-TEL. vol. 13450, pp. 31–44 (2022)
5. Brusilovsky, P., Peylo, C. : Adaptive and Intelligent Web-based Educational Systems. International Journal of Artificial Intelligence in Education (2003)

6. Choffin, B., Popineau, F., Bourda, Y., Vie, J.J. : DAS3H : Modeling Student Learning and Forgetting for Optimally Scheduling Distributed Practice of Skills. In : Junior Conference on Data Science and Engineering (2019)
7. Clément, B., Roy, D., Oudeyer, P.Y., Lopes, M. : Multi-Armed Bandits for Intelligent Tutoring Systems. Journal of Educational Data Mining pp. 20–48 (2015)
8. Corbett, A.T., Anderson, J.R. : Knowledge tracing : Modeling the acquisition of procedural knowledge. In : Proceedings of UMUAI. pp. 253–278 (1994)
9. De Maio, C., Fenza, G., Gaeta, M., Loia, V., Orciuoli, F., Senatore, S. : RSS-based e-learning recommendations exploiting fuzzy FCA for Knowledge Modeling. Applied Soft Computing pp. 113–124 (2012)
10. Ebbinghaus (1885), H. : Memory : A Contribution to Experimental Psychology. Annals of Neurosciences pp. 155–156 (2013)
11. Erdt, M., Fernandez, A., Rensing, C. : Evaluating Recommender Systems for Technology Enhanced Learning : A Quantitative Survey. IEEE Transactions on Learning Technologies pp. 326–344 (2015)
12. Jiang, W., Pardos, Z.A., Wei, Q. : Goal-based Course Recommendation. In : LAK 9th. pp. 36–45 (2019)
13. Nikolayeva, I., Martin, B., Yessad, A., Chenevotot, F., Pilet, J., Prévit, D., Grugeon-Allys, B., Luengo, V. : How to Help Teachers Adapt to Learners ? Teachers' Perspective on a Competency and Error-Type Centered Dashboard. In : 13th EC-TEL. pp. 596–599 (2018)
14. Oudeyer, P.Y., Gottlieb, J., Lopes, M. : Intrinsic motivation, curiosity, and learning. In : Progress in Brain Research, vol. 229, pp. 257–284. Elsevier (2016)
15. Sablayrolles, L., Lefevre, M., Guin, N., Broisin, J. : Design and Evaluation of a Competency-based Recommendation Process. In : Proceedings of ITS (2022)
16. Venant, R., Teyssie, C., Marquié, D., Vidal, P., Broisin, J. : A Competency-Based Model to Bridge the Gap between Academic Trainings and Industrial Trades. In : 15th International Conference on Advanced Learning Technologies (2015)
17. Venant, R., Sharma, K., Dillenbourg, P., Vidal, P., Broisin, J. : A Study of Learners' Behaviors in Hands-On Learning Situations and Their Correlation with Academic Performance. In : AIED. vol. 10331 (2017)
18. Vie, J.J., Kashima, H. : Knowledge Tracing Machines : Factorization Machines for Knowledge Tracing. In : Proceedings of the AAAI Conference on AI (2019)
19. Vlach, H.A., Sandhofer, C.M. : Distributing Learning Over Time : The Spacing Effect in Children's Acquisition and Generalization of Science Concepts. Child Development pp. 1137–1144 (2012)
20. Vygotskiĭ, L.S., Cole, M. : Mind in Society : The Development of Higher Psychological Processes. Harvard University Press (1978)
21. Yanes, N., Mohamed Mostafa, A., Ezz, M., Naif Almuayqil, S. : A Machine Learning-Based Recommender System for Improving Students Learning Experiences. IEEE Access **8**, 201218–201235 (2020)
22. Zhao, Z., Hong, L., Wei, L., Chen, J., Nath, A., Andrews, S., Kumthekar, A., Sathiamoorthy, M., Yi, X., Chi, E. : Recommending what video to watch next : A multitask ranking system. In : Proceedings of the 13th ACM Conference on Recommender Systems. pp. 43–51 (2019)
23. Ziarani, R.J., Ravanmehr, R. : Serendipity in Recommender Systems : A Systematic Literature Review. Journal of Computer Science and Technology (2021)

Proposition d'Indicateurs d'Écriture Collaborative et leur Évaluation

Fahima Djelil[1][0000−0001−8449−2062], Christian Hoffmann[2][000−0002−0620−3621], Anis M. Haddouche[1][0000−0002−5321−3988], Nadine Mandran[2][000−0002−8660−3827], and Cédric d'Ham[2][0000−0002−7313−7097]

[1] IMT Atlantique, Lab-STICC, UMR CNRS 6285, 29238 Brest, France
{fahima.djelil,mohamed-anis.haddouche}@imt-atlantique.fr
[2] Université Grenoble Alpes, CNRS, Grenoble INP, LIG, 38000 Grenoble, France
{christian.hoffmann, nadine.mandran, cedric.dham}@univ-grenoble-alpes.fr

Résumé. L'écriture collaborative est une activité courante en éducation, faisant l'objet d'un champ de recherche en pleine croissance dans le domaine de l'Apprentissage Collaboratif Assisté par Ordinateur. Dans cet article, nous proposons deux indicateurs permettant de mesurer la contribution des apprenants à la rédaction collaborative d'un texte, appelés "balance de contribution" et "co-écriture". Nous nous intéressons également à la relation existante entre les mesures de ces indicateurs avec des stratégies d'écriture collaborative dites "sommative" et "intégrative". De plus, nous évaluons l'interprétation de ces indicateurs et ces stratégies par des enseignants, en utilisant des documents texte rédigés dans une plateforme numérique collaborative appelée *LabNbook*. Les résultats montrent que les enseignants ont une bonne interprétation de ces indicateurs et stratégies.
Mots-clé : Indicateurs · Écriture Collaborative · Apprentissage Collaboratif Assisté par Ordinateur · Analyse de l'Apprentissage · LabNbook.

Abstract. Collaborative Writing is a common activity in education, which is a growing research field in Computer-Supported Collaborative Learning (CSCL). In this paper, we propose two indicators that provide measures of student contributions to a text collaborative writing, namely "balance of contribution" and "co-writing". We are also interested in the existing link between these indicators and collaborative writing strategies called "summative" and "integrative". Moreover, we proceed to the evaluation of the interpretation of the indicators and the strategies by teachers, using text documents written in a collaborative digital platform called LabNbook. Results showed that teachers have a good interpretation of these indicators and strategies.

Keywords: Indicators , Collaborative Writing , Computer Supported Collaborative Learning (CSCL), Learning Analytics , LabNbook.

1 Introduction

L'écriture collaborative suscite un intérêt de recherche croissant dans le domaine de l'apprentissage collaboratif assisté par ordinateur [4,26]. Une grande partie des travaux existants vise à comprendre la dynamique de collaboration entre apprenants, en s'appuyant sur l'analyse de traces numériques, ainsi que les interactions entre pairs tels que les échanges de messages écrits ou les conversations orales [26]. En dépit de l'abondance des travaux dans ce domaine, il existe un besoin accru en termes de mesures et de métriques pouvant servir à mieux comprendre l'écriture collaborative [26].

Dans cet article, nous proposons deux nouveaux indicateurs permettant de mesurer l'écriture collaborative : la *balance de contribution* reflétant dans quelle mesure les contributions des apprenants à un texte co-écrit sont équilibrées et la *co-écriture*, reflétant l'intensité des interactions entre apprenants. Leurs calculs reposent sur une mesure de dispersion basée sur la variance [9]. Nous explorons également le lien existant entre ces indicateurs et des stratégies d'écriture collaborative issues de la littérature. Les définitions de ces indicateurs résultent d'un travail antérieur, que nous ne détaillons pas dans cet article [10]. Ce travail est ancré dans un contexte réel et s'appuie sur une plateforme numérique collaborative, appelée *LabNbook* [6]. Par ailleurs, nous évaluons l'interprétation des indicateurs et des stratégies par des enseignants qui ne sont pas des utilisateurs experts de *LabNbook*. Les questions de recherche que nous abordons sont les suivantes : QR1) Quels sont les indicateurs (métriques) permettant de mesurer la contribution des apprenants en écriture collaborative; QR2) Comment déduire des stratégies d'écriture collaborative à partir de ces indicateurs; QR3) Dans quelle mesure ces indicateurs et ces stratégies sont-ils interprétables par les enseignants ?

Cet article est organisé comme suit. La section 2 présente un état de l'art sur l'écriture collaborative. La section 3 présente les indicateurs proposés ainsi que les stratégies d'écriture collaborative. La section 4 décrit notre méthode de recherche. La section 5 donne le calcul des indicateurs proposés et leur relation avec les stratégies d'écriture collaborative. Les résultats de leur évaluation sont discutés en Section 6, les conclusions et les perspectives sont données en Section 7.

2 L'écriture Collaborative

L'écriture collaborative fait référence à la production d'un texte par deux ou plusieurs auteurs (co-auteurs) [21]. Elle est définie comme un processus impliquant des interactions substantielles entre des apprenants partageant la prise de décision et les responsabilités afin de produire un document unique [14,21,25]. Les taxonomies et modèles existants [18,14,21] ayant pour but de caractériser l'écriture collaborative, témoignent d'un intérêt de recherche soutenu pour ce domaine. Certains modèles offrent un cadre permettant d'examiner les comportements et les stratégies d'écriture collaborative grâce à l'analyse de traces numériques [17,22,16,1,5,12,25].

On peut citer un modèle très connu dans le domaine de langue étrangère [21], fondé sur deux concepts, *l'égalité*, reflétant le niveau de contribution et de contrôle de l'apprenant sur une tâche donnée, et *la mutualité*, reflétant le niveau d'engagement d'un

apprenant sur la contribution d'un autre apprenant. Ces deux fondements s'apparentent à d'autres concepts définis dans des travaux antérieurs. À titre d'exemple, [7] soutient que la collaboration implique la négociation et met l'accent sur le degré de symétrie des interactions entre pairs. Le terme symétrie est emprunté au domaine des systèmes homme-machine et désigne une balance équilibrée de contrôle, induite par la collaboration entre un système et son utilisateur [8]. L'ajustement mutuel est une deuxième caractéristique de la collaboration par la négociation [3]. Il s'agit de stratégies mises en œuvre pour parvenir à un accord dans l'interaction [8]. Cela se reflète par les différentes façons d'éditer un texte écrit par d'autres individus et les façons d'interagir avec ceux qui éditent son propre texte [11]. Cela représente également le degré d'engagement avec les idées et les textes écrits par les autres et la participation à la production d'un écrit commun [13].

Ces travaux nous amènent à penser que l'écriture collaborative peut être caractérisée par deux types d'indicateurs. Le premier mesurant l'égalité des contributions des apprenants, et le second, le degré de leurs interactions.

3 Indicateurs et Stratégies d'Écriture Collaborative

3.1 La Balance de Contribution et la Co-écriture

Nous définissons la *balance de contribution* comme une métrique qui mesure combien les contributions d'apprenants à un texte co-écrit sont égales, équilibrées ou déséquilibrées. Elle s'aligne sur celle proposée par [16], appelée *équilibre de participation*, mesurée en considérant 1 moins la variance des parts des membres d'un groupe dans le travail collaboratif. Dans ce travail [16], le choix de la variance a été justifié par sa simplicité et sa facilité d'interprétation comparée à d'autres méthodes de calcul tel que le coefficient de Gini ou l'indice de Blau. L'indicateur *balance de contribution* est calculé en se basant sur la métrique de variance, mesurant une distribution entre les contributions moyennes des auteurs à la rédaction d'un document en terme de mots.

L'indicateur de *co-écriture* mesure la contribution de l'apprenant, de la même manière que pour la balance de contribution, mais au niveau de la phrase. En effet, nous considérons qu'une phrase est porteuse d'une idée, et l'évaluation de la collaboration dans la rédaction des phrases nous renseigne sur la manière dont les apprenants interagissent lors de l'écriture collaborative. Alors que la balance de contribution mesure la division du travail (nombre de mots qu'un apprenant apporte en contribution à un texte), la co-écriture va au-delà en mesurant la co-construction du texte par les apprenants (modification et ajout de mots dans les phrases écrites par d'autres dans le texte). Les valeurs des deux indicateurs se situent entre 0 et 1 (voir section 5).

3.2 Les Stratégies Sommative et Intégrative

Les indicateurs proposés peuvent nous renseigner sur les stratégies d'écriture collaborative mises en œuvre par un groupe d'apprenants lors de la production d'un texte. Nous nous intéressons particulièrement aux deux stratégies parmi cinq proposées par [17], à savoir :

> *i) construction séquentielle sommative de textes*, *i.e.* un membre du groupe présente un document qui constitue une proposition initiale, partielle ou complète pour la résolution d'une tâche, et les autres participants ajoutent successivement leurs contributions au document initial, sans modifier ce qui a été écrit précédemment, donc en acceptant systématiquement ce qui est ajouté par les autres co-auteurs.
> *ii) construction séquentielle intégrative de textes*, *i.e.* un membre du groupe présente un document qui constitue une proposition de tâches initiales, partielle ou complète, et les autres membres du groupe contribuent successivement à ce document initial, en proposant des modifications justifiées ou en discutant de leur accord ou non avec ce qui a été écrit précédemment.

Nous utilisons le terme *sommative* pour désigner une stratégie selon laquelle chaque apprenant ajoute du texte sans modifier le texte d'autres apprenants, et selon laquelle le résultat est une juxtaposition des contributions individuelles; et le terme *intégrative*, pour désigner une stratégie selon laquelle un apprenant propose une version initiale d'un texte et les autres apprenants contribuent successivement en apportant des modifications au texte existant [10]. Les apprenants n'adoptent pas nécessairement une stratégie unique mais une combinaison des deux.

4 Méthode de Recherche

Notre objectif consiste, d'une part, à construire des indicateurs du niveau de collaboration des apprenants, en s'appuyant sur l'analyse de traces numériques recueillies lors de la production d'un texte par les apprenants, et d'autre part, à vérifier que ces indicateurs sont compréhensibles par les enseignants. Pour ce faire, nous adoptons la méthode de posture épistémologique du Constructivisme Pragmatique (CP) [2], dont la construction des hypothèses de la connaissance scientifique est basée sur la prise en compte de la réalité et de l'humain dans cette réalité. Nous avons choisi la méthode CP car dans ce travail, nous avons besoin de données de terrain produites par les opinions des enseignants pour évaluer les métriques proposées. Nous nous appuyons également sur la méthode Recherche Orientée par la Conception (Design-Based Research) [24] ainsi que les guides associés [15] pour mener notre recherche. Cette méthode propose, entre autres, de construire les connaissances et les outils associés de manière itérative en intégrant les acteurs du terrain.

4.1 Contexte

En vue d'évaluer nos indicateurs, nous utilisons *LabNbook*, une plateforme numérique pour l'apprentissage des sciences expérimentales [6]. Elle fournit des outils pour la rédaction de documents scientifiques collaboratifs et permet aux apprenants d'interagir entre eux. De plus, *LabNbook* fonctionne en mode "co-édition verrouillée" [23], c'est-à-dire que les apprenants peuvent travailler simultanément dans l'espace de travail partagé mais chaque document (appelé *LabDoc*) qui le compose ne peut être édité que par un seul apprenant à la fois.

4.2 Collecte de Données

Afin d'évaluer dans quelle mesure ces stratégies et ces indicateurs sont interprétables par les enseignants, nous avons administré un questionnaire à 15 enseignants, utilisateurs non-experts de la plateforme *LabNbook*. Nous avons sélectionné 12 LabDocs, chacun est co-écrit par 2 à 4 élèves dans des conditions réelles (activités en classe). Les contributions à l'écriture des différents apprenants ont été mises en évidence. Nous avons choisi des LabDocs ayant différentes stratégies d'écriture : 4 LabDocs selon une stratégie intégrative, 4 LabDocs selon une stratégie sommative, et 4 LabDocs selon des stratégies mixtes. Les 15 enseignants ont examiné les 12 LabDocs, un à la suite de l'autre. Après avoir lu les définitions des stratégies et des indicateurs et visualisé chaque LabDoc, ils indiquent :

(1) la stratégie d'écriture parmi 6 choix : EI (Entièrement Intégrative), PI (Plutôt Integrative), ESI (Entre Sommative et Intégrative), PS (Plutôt Sommative), ES (Entièrement Sommative) ou Je ne sais pas ;
(2) une estimation du niveau des indicateurs, parmi 3 choix : Faible (F), Moyen (M) ou Élevé (E) ;
(3) une valeur numérique pour chaque indicateur entre 0 et 1.

5 Calcul des Indicateurs

5.1 Comparaison de Séquences de Textes

Afin de construire nos indicateurs, nous examinons d'abord l'évolution d'un texte afin de la quantifier. Pour ce faire, nous comparons des séquences de textes par paires, en utilisant la méthode dite *Sequence Matcher* dérivant de l'algorithme *Gestalt Pattern Matching* [19] [3]. Cette méthode examine l'évolution d'un texte entre deux versions. Cela consiste tout d'abord à repérer la plus longue (en termes de nombre de caractères) séquence contiguë d'appariement qui ne contient aucun élément dit "inutile", tels que les lignes vides ou les espaces blancs. Ensuite, cette opération est répétée pour les séquences se trouvant à gauche et à droite de cette séquence contiguë. Enfin, afin de nommer les changements dans le texte, une étiquette est associée à chaque séquence : *Equal* (les séquences sont égales), *Insert* (la séquence est insérée), *Delete* (la séquence est supprimée), *Replace* (la séquence est remplacée).

5.2 Matrice de Contribution

Nous désignons par *matrice de contribution*, une matrice mathématique qui fournit pour chaque mot d'un texte final, le niveau de contribution de chacun des apprenants ayant co-écrit ce texte. Dans cette matrice, les lignes représentent les apprenants contributeurs (co-auteurs) et les colonnes représentent les mots constituant le texte final. C'est une matrice $K \times M$, où K est le nombre d'auteurs et M le nombre de mots co-écrits par les K auteurs. Dans cette matrice, le niveau de contribution d'un auteur i à un mot l dans une

[3] *Bibliothèque Difflib https://github.com/python/cpython/blob/main/Lib/*difflib.*py*

phrase j est définit par $x_{i,j,l} \in [0,1]$, où $i \in [1,K]$, $j \in [1,N]$ et $l \in [1,n_j]$, N est le nombre de phrases composant le texte, et n_j désigne le nombre de mots dans la phrase j. Il est égal à 1 lorsque le mot est entièrement écrit par cet auteur, et à 0, lorsque l'auteur ne contribue pas à l'écriture de ce mot. La matrice de contribution ne prend pas en compte les mots supprimés. Le total des contributions des auteurs à un mot est égal à 1.

5.3 Balance de Contribution

La balance de contribution est un indicateur qui mesure combien les contributions d'écriture collaborative des apprenants sont égales (bien équilibrées) ou non en termes de mots constituant le texte. Plus la valeur de cet indicateur est proche de 1, plus les contributions des apprenants sont égales ou équilibrées. À l'inverse, plus il est proche de 0, moins les contributions des apprenants sont égales ou équilibrées. Cet indicateur est basé sur la variance des contributions moyennes des auteurs, reflétant la distance entre la contribution moyenne de chaque auteur ($\overline{x}_{i,\cdot,\cdot}$) et les contributions moyennes de l'ensemble des auteurs aux mots composant le texte. Notons que, si les auteurs contribuent de manière équilibrée, alors leurs scores sont proches de la moyenne et ils contribuent de manière tout à fait égale. De plus, afin de pénaliser le cas où le texte entier est écrit par un seul auteur, cette mesure est normalisée. Enfin, par souci de facilité d'interprétation pour les enseignants, cet indicateur est donné par 1 moins une valeur de dispersion, ce qui donne l'indicateur balance de contribution comme suit

$$e(X) = 1 - \frac{K}{K-1} \sum_{i=1}^{K} \left(\overline{x}_{i,\cdot,\cdot} - \frac{1}{K} \right)^2 . \tag{1}$$

Notons que, $\sum_{i=1}^{K} \overline{x}_{i,\cdot,\cdot} = 1$ et que la moyenne des contributions moyennes de l'ensemble des auteurs est donnée par

$$\overline{x}_{\cdot,\cdot,\cdot} = \frac{1}{K} \sum_{i=1}^{K} \overline{x}_{i,\cdot,\cdot} = \frac{1}{K} .$$

Cet indicateur (1) atteint sa valeur maximale 1, lorsque la valeur de la dispersion soustraite à 1 atteint sa valeur minimale 0, autrement dit, lorsque tous les auteurs contribuent de manière égale (ou équilibrée) au texte, c'est-à-dire lorsque la contribution moyenne de chaque auteur $\overline{x}_{i,\cdot,\cdot}$ est égale à $1/K$. Elle atteint sa valeur minimale 0 lorsqu'un auteur écrit seul la totalité du texte.

5.4 La Co-écriture

À l'instar de la balance de contribution, mesurant les contributions des auteurs en terme de mots à un texte, l'indicateur de co-écriture mesure les contributions des auteurs au niveau des phrases. Ainsi, la construction de cet indicateur nécessite en amont le découpage du texte en phrases. Pour ce faire, nous utilisons une approche de segmentation de phrases, basée sur des règles (heuristiques) pour sa simplicité de mise en œuvre [20]. Nous utilisons un outil de détection des limites de phrases basé sur un ensemble de

règles appelé *Golden Rule Set* [4], conçues à la main permettant de déterminer les limites de phrases, telles que les ponctuations.

Grâce à la matrice de contribution, nous calculons pour chaque auteur sa contribution moyenne à l'ensemble des phrases. Nous pénalisons le cas où un auteur écrit seul une phrase, et nous calculons pour une phrase j, 1 moins une valeur de dispersion, comme suit

$$e_j(X) = 1 - \frac{K}{K-1} \sum_{i=1}^{K} \left(\bar{x}_{i,j,\bullet} - \frac{1}{K} \right)^2 .$$

Ainsi, l'indicateur de co-écriture est donné par

$$c(X) = \sum_{j=1}^{N} p_j e_j(X) \quad \text{où} \quad p_j = \frac{n_j}{M} \tag{2}$$

est le poids de la phrase j. Lorsque toutes les phrases sont écrites par un seul auteur, l'indicateur de co-écriture est égal à 0. Il est égal à 1 lorsque toutes les phrases sont co-écrites de manière équilibrée par l'ensemble des auteurs.

Les deux métriques (1) et (2) permettent de répondre à la QR1. Par ailleurs, les deux indicateurs possèdent une propriété découlant de ces métriques. En effet, l'indicateur de co-écriture $c(X)$ dans (2) est inférieur ou égal à l'indicateur balance de contribution $e(X)$ dans (1) [9].

5.5 Relation des Indicateurs aux Stratégies d'Écriture Collaborative

Ces indicateurs ont des valeurs continues comprises entre 0 et 1. Par conséquent, nous pouvons représenter les documents texte co-écrits dans un plan à deux dimensions où l'axe des abscisses correspond à la balance de contribution et l'axe des ordonnées correspond à la co-écriture (Figure 1). De plus, étant donné que l'indicateur de co-écriture est toujours inférieur ou égal à la balance de contribution, l'espace de représentation des documents dans le plan 2D est réduit à un demi plan.

Il est alors possible de distinguer les documents écrits selon des stratégies sommatives (ES, PS) ou intégratives (EI, PI) de ceux écrits selon une stratégie mixte (ESI), grâce à leur emplacement dans ce plan 2D. Cela nous permet de répondre à la QR2. En effet, lorsque la co-écriture est faible, la stratégie est sommative. Par conséquent, les documents écrits selon des stratégies sommatives sont représentés près de l'axe des x. Lorsque les stratégies sont intégratives, les documents sont représentés près de la diagonale du plan 2D et les valeurs des deux indicateurs sont proches. Les documents rédigés selon une stratégie mixte sont représentés au milieu de l'espace entre l'axe des x et la diagonale.

Nous illustrons dans la figure 1 trois exemples de documents, un premier écrit avec une stratégie intégrative (EI) ayant une balance de contribution égale à 0.92 et de co-écriture égale à 0.79, un second document écrit avec une stratégie sommative (ES) ayant comme coordonnées $(0.92, 0.05)$ et un troisième écrit selon une stratégie mixte (ESI) dont les coordonnées sont $(0.77, 0.33)$.

[4] `https://github.com/diasks2/pragmatic_segmenter`

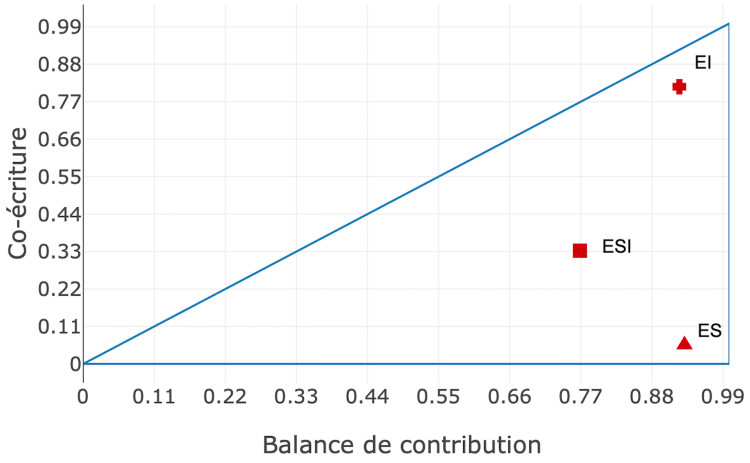


Fig. 1. Plan 2D montrant le lien entre les indicateurs et les stratégies d'écriture collaborative de documents texte.

6 Résultats et Discussion

Les résultats du questionnaire nous ont permis d'évaluer les interprétations des stratégies et des indicateurs par les enseignants (QR3). Nous avons comparé les valeurs et les niveaux des indicateurs estimés par les enseignants aux valeurs machine, i.e. calculées à l'aide des formules (1) et (2). Le niveaux d'un indicateur est faible, si sa valeur se situe dans $[0, 1/3[$, moyen si sa valeur se situe dans $[1/3, 2/3[$ et élevé si sa valeur se situe dans $[2/3, 1]$.

Le tableau 1 compare pour chaque LabDoc et pour chaque stratégie le nombre de réponses reçues (en gras les LabDocs mal classés). On remarque que $9/12$ LabDocs sont bien classés avec une majorité de bonnes réponses variant entre 8 et 14. Les LabDocs 1 et 6 sont mal classés, mais présentent respectivement 6 et 5 bonnes réponses. Ceci est probablement dû aux formules mathématiques présentes dans le LabDoc 1, pouvant influencer la perception des enseignants (les formules mathématiques n'étant pas prises en compte à ce stade), et à la difficulté pour les enseignants de percevoir la stratégie mixte du LabDoc 6. Pour le LabDoc 12, les réponses des enseignants sont plus dispersées par rapport aux LabDocs 1 et 6, mais avec 5 bonnes réponses et 1 réponse "Je ne sais pas". Le LabDoc 12 présente un niveau de collaboration très faible, pouvant conduire à une mauvaise interprétation de la stratégie par les enseignants.

Les enseignants ont donné une estimation des indicateurs, en choisissant un niveau parmi les 3 proposés (faible, moyen et élevé), et une valeur numérique entre 0 et 1. Le tableau 2 fournit pour chaque LabDoc le nombre d'apprenants ayant contribué à sa rédaction, le nombre de réponses des enseignants pour chaque niveau d'indicateur (faible, moyen et élevé), le niveau de chaque indicateur, ainsi que sa valeur machine. Nous avons également calculé l'écart entre la valeur machine (VM) d'un indicateur

Table 1. Vérification des stratégies perçues par les enseignants. En gras on distingue les cas où la majorité des réponses des enseignants ne correspond pas à la stratégies réelle d'un LabDoc.

LabDoc	Stratégie	ES	PS	ESI	PI	EI	Je ne sais pas
1	**EI**	**0**	**2**	**4**	**3**	**6**	**0**
2	ES	11	2	1	1	0	0
3	EI	0	1	1	1	12	0
4	ESI	0	1	8	5	1	0
5	EI	2	0	0	2	11	0
6	**ESI**	**0**	**8**	**5**	**2**	**0**	**0**
7	ES	12	1	0	0	2	0
8	PI	0	0	4	8	3	0
9	PS	0	8	5	1	0	1
10	ES	14	0	1	0	0	0
11	ES	14	1	0	0	0	0
12	**EI**	**1**	**4**	**2**	**2**	**5**	**1**

$\theta \in [0,1]$ et l'estimation des enseignants, en utilisant l'Écart Quadratique Moyen (EQM)

$$EQM = \sqrt{\frac{\sum_1^n (\hat{\theta}_i - \theta)^2}{n}}$$

où, pour $i = 1,\ldots,n$, $\hat{\theta}_i \in [0,1]$ est l'estimation d'un enseignant et $n = 15$ est le nombre d'enseignants ayant répondu au questionnaire.

Les résultats du tableau 2 montrent que $8/12$ LabDocs sont bien classés par les enseignants concernant les niveaux de l'indicateur balance de contribution et $10/12$ concernant les niveaux de l'indicateur de co-écriture. Nous remarquons que l'EQM prend des valeurs dans l'intervalle $[0.07, 0.39]$ concernant la balance de contribution et dans l'intervalle $[0.04, 0.42]$ concernant la co-écriture. Les résultats montrent que lorsque les enseignants parviennent à percevoir un niveau correct pour un indicateur ils parviennent aussi à bien estimer sa valeur. De plus, concernant la balance de contribution, les LabDocs pour lesquels le niveau n'est pas correctement estimé, sont ceux écrits par plus de deux apprenants sauf pour le LabDoc 7. Concernant la co-écriture, le niveau n'est pas correctement estimé pour le LabDoc 1 (probablement à cause de la présence de formules mathématiques) et le LabDoc 5 (probablement à cause de la difficulté de percevoir correctement la co-écriture sur un LabDoc rédigé par plus de deux auteurs, et selon une stratégie intégrative).

A partir de ces résultats, et en réponse à la QR3, nous pouvons déduire que les indicateurs et les stratégies sont bien interprétés par les enseignants dans une certaine mesure. Cependant, les stratégies des LabDocs comprenant des formules mathématiques, ou rédigés selon des stratégies mixtes ou présentant un très faible niveau de collaboration peuvent être difficiles à percevoir. En ce qui concerne les indicateurs, des difficultés dans leur estimation sont également observées pour les LabDocs écrits par plus de 2 co-auteurs.

Table 2. Comparaison des niveaux d'indicateurs perçus Faible (F), Moyen (M) et Élevé (E) et des valeurs estimées par les enseignants avec les niveaux (NM) et les valeurs machines (VM). En gras on distingue les cas où la majorité des réponses des enseignants ne correspondent pas aux résultats des mesures calculées par la machine. NB désigne le nombre de co-auteurs d'un LabDoc et EQM l'écart quadratique moyen.

		Balance de contribution						Co-écriture					
LabDoc	NB	F	M	E	NM	VM	EQM	F	M	E	NM	VM	EQM
1	2	0	8	7	M	0.49	0.32	**0**	**4**	**11**	**M**	**0.41**	**0.42**
2	2	0	6	9	E	0.87	0.15	13	1	1	F	0.19	0.11
3	2	1	2	12	E	0.84	0.24	1	0	14	E	0.76	0.28
4	2	0	1	14	E	0.99	0.21	3	9	3	M	0.41	0.22
5	3	**15**	**0**	**0**	**M**	**0.49**	**0.36**	**12**	**1**	**2**	**M**	**0.44**	**0.35**
6	2	0	1	14	E	0.99	0.19	7	7	1	M	0.50	0.2
7	2	**13**	**2**	**0**	**M**	**0.51**	**0.3**	14	1	0	F	0.07	0.09
8	2	0	7	8	E	0.99	0.3	0	6	9	E	0.74	0.21
9	4	**13**	**2**	**0**	**M**	**0.48**	**0.27**	13	2	0	F	0.17	0.12
10	3	**1**	**11**	**3**	**E**	**0.92**	**0.39**	13	2	0	F	0.09	0.1
11	2	15	0	0	F	0.04	0.04	15	0	0	F	0.29	0.19
12	2	15	0	0	F	0.14	0.07	14	1	0	F	0.09	0.09

7 Conclusion

Ce travail contribue à la recherche sur l'écriture collaborative, en proposant deux nouveaux indicateurs. Le premier permet de mesurer l'égalité des contributions en terme de mots composant un texte (balance des contributions), et le second mesure les interactions entre apprenants en termes des contributions au niveau des phrases (co-écriture).

Une deuxième contribution de l'article consiste à identifier les stratégies d'écriture collaborative sommative et intégrative qui sont bien définies dans la littérature et utiles dans un contexte pédagogique, par exemple pour distinguer les travaux coopératifs (sommatifs) des travaux collaboratifs (intégratifs) [17]. Nous avons montré que ces stratégies d'écriture peuvent être dérivées des deux indicateurs proposés.

Une troisième contribution de cet article consiste en l'évaluation de l'interprétation des indicateurs et des stratégies proposées. Nous avons utilisé des documents texte édités dans la plateforme *LabNbook* et issus d'un contexte réel. Nous avons réalisé une évaluation par questionnaire auprès d'enseignants qui ne sont pas des utilisateurs experts de *LabNbook*. Les résultats du questionnaire ont permis de comparer la perception des indicateurs par les enseignants avec les résultats des calculs automatiques, et de vérifier les stratégies perçues. Les résultats ont montré que les enseignants ont une bonne interprétation des indicateurs et des stratégies. En effet, les enseignants étaient capables d'estimer et de percevoir correctement les indicateurs et les stratégies de co-écriture de la plupart des documents.

Enfin, nous concluons que les nouvelles métriques proposées pour la balance de contribution et la co-écriture peuvent aider à évaluer le travail collaboratif d'apprenants, en caractérisant le niveau de collaboration et les stratégies d'écriture collaborative. En

perspective, il serait intéressant d'étudier comment permettre aux enseignants d'exploiter ces indicateurs à partir de visualisations appropriées, et d'évaluer leur acceptabilité, utilisabilité et utilité. De plus, ce travail n'est pas sans limites. Il serait pertinent d'améliorer notre approche d'analyse de texte, afin de prendre en compte les formules mathématiques. En effet, notre méthode de comparaison de séquences de textes peut produire des résultats erronés lorsque les documents texte comportent des formules mathématiques. Une amélioration possible est de repérer les formules mathématiques en utilisant les techniques d'apprentissage automatique, comme les Chaîne de Markov Cachées.

Remerciement. Le projet est co-financé par la Région Bretagne dans le cadre de son programme Stratégie d'Attractivité Durable (SAD) et le Conseil départemental du Finistère dans le cadre de son dispositif Aide aux Programmes de Recherche Emergents (APRE) (2021-2023). Les auteurs tiennent à remercier l'ensemble des enseignants ayant contribué à cette étude.

Références

1. Abrams, Z. : Exploring collaboratively written l2 texts among first-year learners of german in google docs. Computer Assisted Language Learning **29**(8), 1259–1270 (2016)
2. Avenier, M.J., Thomas, C. : Finding one's way around various methodological guidelines for doing rigorous case studies : A comparison of four epistemological frameworks. Systemes d'information management **20**(1), 61–98 (2015)
3. Baker, M. : A model for negotiation in teaching-learning dialogues. Journal of Interactive Learning Research **5**(2), 199 (1994)
4. Chen, J., Wang, M., Kirschner, P.A., Tsai, C.C. : The role of collaboration, computer use, learning environments, and supporting strategies in cscl : A meta-analysis. Review of Educational Research **88**(6), 799–843 (2018)
5. Cho, H. : Synchronous web-based collaborative writing : Factors mediating interaction among second-language writers. Journal of Second Language Writing **36**, 37–51 (2017)
6. d'Ham, C., Wajeman, C., Girault, I., Marzin-Janvier, P. : LabNbook, plateforme numérique support des pédagogies actives et collaboratives en sciences expérimentales. In : Broisin, J., Sanchez, E., Yessad, A., Chenevotot, F. (eds.) EIAH 2019 : Environnement Informatiques pour l'Apprentissage Humain. pp. 49–60. Actes de la 9ème Conférence sur les Environnements Informatiques pour l'Apprentissage Humain, Paris, France (Jun 2019)
7. Dillenbourg, P. : What do you mean by collaborative learning ?, p. 1–19. Elsevier, Oxford (1999)
8. Dillenbourg, P., Michael, B. : Negotiation spaces in human-computer collaborative learning. In : Proceedings of the International Conference on Cooperative Systems. p. 12–14 (1996)
9. Haddouche, M.A., Djelil, F., Hoffmann, C., Mandran, N., d'Ham Cédric. : Proposal of indicators for measuring collaborative writing in a digital learning environment. 15th International Conference on Computer Supported Education **2**, 495–502 (2023)
10. Hoffmann, C., Mandran, N., d'Ham, C., Rebaudo, S., Haddouche, M.A. : Development of actionable insights for regulating students' collaborative writing of scientific texts. In : Educating for a New Future : Making Sense of Technology-Enhanced Learning Adoption. p. 534–541. Lecture Notes in Computer Science, Springer International Publishing, Cham (2022)

11. Larsen-Ledet, I., Korsgaard, H. : Territorial functioning in collaborative writing. Computer Supported Cooperative Work (CSCW) **28**(3), 391–433 (Jun 2019)
12. Li, M., Kim, D. : One wiki, two groups : Dynamic interactions across esl collaborative writing tasks. Journal of second language writing **31**, 25–42 (2016)
13. Li, M., Zhu, W. : Explaining dynamic interactions in wiki-based collaborative writing. Language Learning & Technology **21**(2), 96–120 (2016)
14. Lowry, P.B., Curtis, A., Lowry, M.R. : Building a taxonomy and nomenclature of collaborative writing to improve interdisciplinary research and practice. The Journal of Business Communication (1973) **41**(1), 66–99 (2004)
15. Mandran, N., Vermeulen, M., Prior, E. : Thedre's framework : Empowering phd candidates to efficiently implement design-based research. Education and Information Technologies pp. 1–24 (2022)
16. Olson, J.S., Wang, D., Olson, G.M., Zhang, J. : How people write together now : Beginning the investigation with advanced undergraduates in a project course. ACM Transactions on Computer-Human Interaction **24**(1), 4 :1–4 :40 (Mar 2017)
17. Onrubia, J., Engel, A. : Strategies for collaborative writing and phases of knowledge construction in cscl environments. Computers & Education **53**(4), 1256–1265 (2009)
18. Posner, I.R., Baecker, R.M. : How people write together (groupware). In : Proceedings of the Twenty-Fifth Hawaii International Conference on System Sciences. vol. 4, p. 127–138. IEEE (1992)
19. Ratcliff, J.W., Metzener, D.E. : Pattern-matching-the gestalt approach. Dr Dobbs Journal **13**(7), 46 (1988)
20. Sadvilkar, N., Neumann, M. : PySBD : Pragmatic sentence boundary disambiguation. In : Proceedings of Second Workshop for NLP Open Source Software (NLP-OSS). pp. 110–114. Association for Computational Linguistics, Online (Nov 2020)
21. Storch, N. : Collaborative Writing in L2 Classrooms. Multilingual Matters, Bristol, Blue Ridge Summit (2013)
22. Sundgren, M., Jaldemark, J. : Visualizing online collaborative writing strategies in higher education group assignments. The International Journal of Information and Learning Technology **37**(5), 351–373 (Aug 2020)
23. Wang, D., Tan, H., Lu, T. : Why users do not want to write together when they are writing together : Users' rationales for today's collaborative writing practices. Proceedings of the ACM on Human-Computer Interaction **1**(CSCW), 1–18 (2017)
24. Wang, F., Hannafin, M.J. : Design-based research and technology-enhanced learning environments. Educational technology research and development **53**(4), 5–23 (2005)
25. Zhang, M., Chen, W. : Assessing collaborative writing in the digital age : An exploratory study. Journal of Second Language Writing **57**, 100868 (sep 2022)
26. Zhang, M., Gibbons, J., Li, M. : Computer-mediated collaborative writing in l2 classrooms : A systematic review. Journal of Second Language Writing **54**, 100854 (2021)

Usages dans le temps d'un tableau de bord d'apprentissage dans un jeu sérieux

Katia Oliver-Quelennec[1,3][0000−0002−7318−7449], François Bouchet[1][0000−0001−9436−1250], Thibault Carron[1][0000−0001−6982−7055] et Claire Pinçon[2][0000−0002−5509−6199]

[1] Sorbonne Université, CNRS, LIP6, F-75005 Paris, France
{francois.bouchet,thibault.carron}@lip6.fr
[2] Univ. Lille, CHU Lille, ULR 2694 - METRICS : Évaluation des technologies de santé et des pratiques médicales, F-59000 Lille, France
[3] Univ. Lille, GIVRE, DAPI, France
{katia.quelennec,claire.pincon}@univ-lille.fr

Résumé. Les tableaux de bord d'apprentissage (TBA) sont des outils pour l'enseignement aux nombreux atouts. Notre étude s'intéresse à leur adaptation dans le temps, dans le cadre d'un jeu sérieux sur la pratique officinale. L'approche générique suivie consiste à (1) observer l'usage d'un TBA co-conçu et (2) questionner les étudiants sur celui-ci et sur leurs attentes dans le temps. A partir de ce jeu de données ($N = 77$ réponses et $N = 121$ traces), nous analysons les évolutions des attentes et usages selon le planning pédagogique, et confrontons ces sources d'information entre elles. Le TBA proposé a été évalué positivement par la majorité des répondants. De légères variations dans les attentes des étudiants et leurs consultations des différentes pages du TBA selon le moment du semestre apparaissent, plaidant en faveur d'une adaptation dans le temps. Ce travail montre l'importance de croiser traces d'apprentissage et avis des utilisateurs pour identifier leurs besoins, et ouvrent des perspectives de modèles d'adaptation automatiques.

Mots clefs : Tableau de bord d'apprentissage , Adaptation , Jeu sérieux

Abstract. Learning Analytics Dashboards (LADs) are teaching tools with many strengths. Our study focuses on their adaptation over time in a serious game on pharmacy practice. The generic approach taken is to (1) observe the use of a collaboratively designed LAD and (2) survey students about this tool and their expectations over time. Using this dataset (77 responses and 121 traces), we analyze the evolution of expectations and uses according to the pedagogical planning and compare these sources of information. The proposed LAD was positively evaluated by most respondents, and slight differences in students' expectations and their use of the different LAD pages according to the semester time emerge, suggesting adaptation over time is needed. This work shows the importance of intersecting learning traces and user opinions to determine their needs, and opens perspectives for automatic adaptation models.

Mots-clés : Learning dashboard , Adaptation , Serious game

1 Travaux antérieurs et questions de recherche

Dans les Environnements Informatiques pour l'Apprentissage Humain (EIAH), le tableau de bord d'apprentissage (TBA) est un outil en plein essor pouvant apporter un soutien à l'apprentissage [2] mais nécessitant d'être adapté au contexte d'apprentissage [2], adaptable par l'utilisateur [6] et adaptatif par le système [1]. Si les travaux de recherche autour de l'adaptation s'intéressent beaucoup à la question de l'adaptation à des profils d'apprenants différents, l'adaptation dans le temps semble peu explorée. Pour Verbert et al. [7], les TBA "devraient également réagir différemment selon les étudiants, à différents moments du cours", et ils soulignent l'importance de la prise en compte du moment dans le semestre et de l'évolution de l'étudiant. Pour Park et al. [4], "la conception du TBA et des instructions devrait prendre en compte les objectifs d'apprentissage des étudiants selon des temps spécifiques [...] afin de fournir un meilleur soutien tout au long du processus d'apprentissage", et ils proposent donc une personnalisation du TBA en couplant le moment dans le semestre et les objectifs d'apprentissage propres à chaque étudiant.

Notre contribution scientifique vise à développer cette approche, en analysant l'intérêt d'une adaptation temporelle pour l'apprenant en termes d'utilisabilité ou d'apprentissage, et en identifiant les adaptations temporelles éventuellement souhaitées par les étudiants. Ainsi, nous essayons de répondre à trois sous-questions (SQ) autour de la perception et l'usage d'un TBA dans le temps :

— SQ1 : quelles sont les différences de besoins de TBA exprimées par des étudiants utilisateurs d'un TBA au cours d'un semestre ?

— SQ2 : est-ce que les usages d'un TBA évoluent le temps d'un semestre ?

— SQ3 : y a-t-il un accord entre attentes et usages pour des TBA différents dans le temps ?

2 Matériel et méthode

Notre étude s'appuie sur le TBA d'un jeu sérieux numérique utilisé en complément de mises en situation en présentiel pour développer les compétences professionnelles des futurs pharmaciens. Une étude sur une première version du TBA a permis d'identifier un impact positif sur l'apprentissage pour ceux qui utilisaient le TBA [3]. Le TBA actuel, co-conçu avec les étudiants (cf. figure 1), est composé plusieurs pages liées à différents objectifs : (1) page **planification** (cadre bleu foncé dans 1) regroupant emploi du temps des enseignements dirigés de l'unité d'enseignement, date de l'examen, nombre de cas cliniques disponibles par pathologie, outil pour planifier le nombre de cas à faire par semaine pour terminer le jeu à une date choisie et suivi du nombre de cas faits par mois, (2) page **communication** composée de liens vers les forums du jeu (pédagogique et technique) et un chat dédié, (3) page **évaluation** (cadre vert dans 1) avec un diagramme en radar avec le pourcentage de réussite obtenu dans chaque pathologie et le pourcentage de réussite attendu par les enseignants (fixé à 60%),

(4) page **évolution en mode challenge** (cadre rose dans 1) pour les étudiants souhaitant partager et comparer leurs résultats par pathologie entre pairs, (5) page **évolution en mode révision** (cadre bleu clair dans 1) listant chaque cas clinique déjà fait avec les résultats obtenus et un lien direct vers les cas.

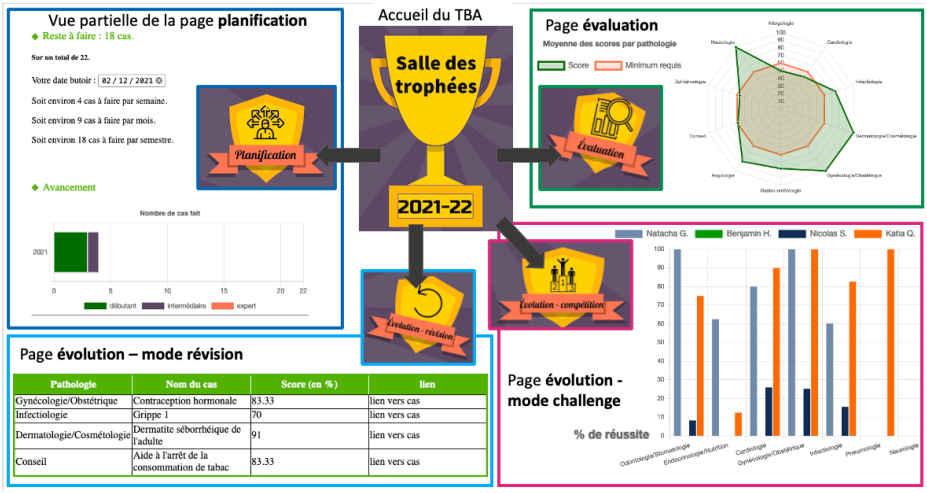


FIG. 1 – Exemples de pages du TBA du jeu

Matériel. Nous avons conçu un **questionnaire** visant à (1) évaluer ce nouveau TBA et (2) avoir l'avis des étudiants sur l'utilité de chaque page selon le moment du semestre. Il inclut des questions sur le profil des étudiants, leur utilisation du TBA, l'utilité des pages selon le moment du semestre et l'évaluation du TBA selon les critères de Park et al. [5]. Les étudiants pouvaient saisir leurs noms afin de confronter leurs avis à leur usage réel du TBA. Nous analysons aussi les **traces des activités numériques** via les logs de la plateforme de formation donnant accès au jeu (Moodle). Seul l'accès au jeu étant tracé par défaut, nous avons donc tracé séparément les activités de chaque tentative des étudiants (numéro de la tentative, dates de début et de fin).

Méthode. Différents prétraitements ont été effectués pour éviter d'éventuels biais. Nous avons supprimé les traces de la première consultation de chaque page du TBA, considérant qu'elles correspondaient plus à un temps de découverte qu'à un véritable usage. Pour observer si les usages du TBA évoluent selon le moment du semestre, nous avons fait des analyses sur toute la cohorte, en comparant les proportions d'utilisateurs ayant consulté les pages du TBA selon les mois. Suite à une limite technique de Moodle qui ne trace que l'entrée dans le module, nous avons considéré uniquement les tentatives de jeu d'une courte durée (< 29 jours, 70% des tentatives) pour identifier le mois de consultation des pages du TBA.

3 Résultats

Questionnaire. De janvier à mai 2022, nous avons obtenu $N = 121$ réponses complètes (sur 199 étudiants) dont 77 avec identification permettant de croiser

leurs avis avec leurs traces. 40 étudiants (sur 121) ont déclaré avoir utilisé le TBA et en ont fait une évaluation plutôt positive. Parmi les deux tiers de non-utilisateurs déclarés (81 étudiants), 35% d'étudiants qui ne l'ont pas identifié ou ne savaient pas où le consulter n'ont effectivement pratiquement aucune trace sur le TBA, mais les 65% qui l'ont estimé inutile l'ont pratiquement tous visité à plusieurs reprises, jusqu'à plus de 10 fois pour un tiers d'entre eux.

Évolution des attentes au cours du semestre. Pour répondre à **SQ1**, le questionnaire nous a permis de recueillir les avis sur l'utilité de chaque page du TBA, à choisir entre début, milieu, fin, tout au long du semestre ou pas du tout d'utilité. Selon le planning pédagogique, le début de semestre correspond au mois d'octobre, le milieu à novembre et la fin à décembre (mois d'évaluation). Les résultats des 40 utilisateurs déclarés sont présentés en figure 2a. Les trois quarts d'entre eux ont estimé qu'au moins une page était plus utile à un moment différent du semestre. En excluant les pages jugées inutiles ou utiles tout au long du semestre, nous pouvons identifier de légères variations : la planification baisse dans le semestre, la communication, les pages évolutions challenge et révision montent puis baissent en fin de semestre, tandis que les besoins de la page évaluation augmentent avec le temps.

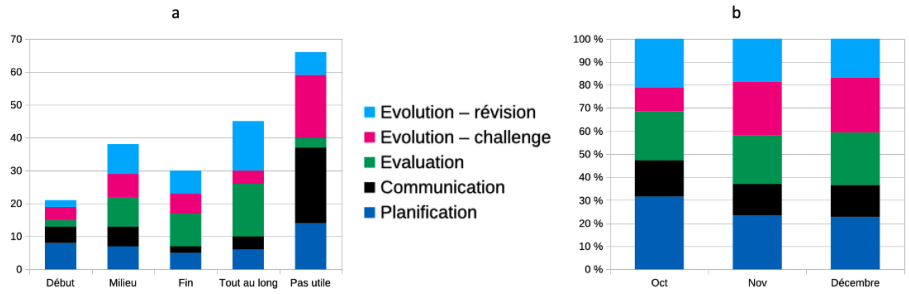


FIG. 2 – Selon le moment du semestre : perception de l'utilité des pages TBA (a) et évolution de la proportion des consultations (b)

Évolution des usages au cours du semestre. Pour répondre à **SQ2**, nous avons comparé les proportions de consultations de chaque page par rapport aux autres parmi les utilisateurs du TBA. Les résultats sont présentés dans la figure 2b. Là aussi, des variations modérées apparaissent : la consultation des pages planification et évolution mode révision baisse, celle des pages communication et évaluation est stable tout au long du semestre, et les consultations de la page évaluation en mode challenge augmentent dans le temps.

Accord entre utilité déclarée et usage réel. Pour répondre à **SQ3**, nous confrontons les tendances d'évolution des perceptions de l'utilité des pages du TBA d'après les étudiants et les traces recueillies (cf. tableau 1 où une flèche illustre l'évolution entre chaque moment du semestre), et observons quelques différences : (1) **Planification** : l'avis et le comportement sont plutôt cohérents en considérant cette page plus utile en début de semestre puis de moins en moins utile. Ce retour semble logique car la planification fait sens en début et milieu de

semestre plutôt qu'en fin de semestre. (2) **Communication** : cette page est jugée plus utile en milieu de semestre avant de rebaisser en fin de semestre. D'après les traces, sa consultation baisse légèrement puis se stabilise en fin du semestre. Les liens vers les forums du jeu sont toujours utiles car c'est le moyen d'avoir des réponses à d'éventuelles questions, pouvant arriver à tout moment dans le semestre. (3) **Évaluation** : cette page est jugée plutôt utile en fin de semestre, mais les traces montrent une consultation stable au long du semestre avec une légère augmentation à la fin. L'impression des étudiants est peut-être qu'il est plus utile de s'évaluer en fin de semestre avant l'examen, mais il peut aussi être intéressant de suivre son niveau tout au long du semestre (ex : pour voir s'il est utile de continuer d'utiliser le module). (4) **Évolution mode challenge** : cette page est jugée plus utile en milieu de semestre et selon les traces, sa consultation augmente en milieu de semestre puis se stabilise. Nous pouvons imaginer qu'il est plus motivant de travailler en groupe. (5) **Évolution mode révision** : cette page est jugée utile principalement en milieu de semestre et les traces montrent un usage qui décroît faiblement entre le début et la fin du semestre. Ce résultat est contre-intuitif car nous pourrions imaginer que cette page serait plus utile pour préparer l'évaluation de l'UE.

TABLE 1 – Évolution de la perception de l'utilité des pages du TBA d'après les étudiants et évolution de la proportion d'utilisation des pages du TBA

Page TBA	Perception	Traces
Planification	$\downarrow\downarrow$	$\downarrow\rightarrow$
Communication	$\uparrow\downarrow$	$\downarrow\rightarrow$
Évaluation	$\uparrow\uparrow$	$\rightarrow\uparrow$
Évolution mode challenge	$\uparrow\downarrow$	$\uparrow\rightarrow$
Évolution mode révision	$\uparrow\downarrow$	$\downarrow\downarrow$

4 Discussion et conclusion

Discussion. Que ce soit de l'avis des étudiants ou d'après les traces d'utilisation, les variations dans le temps restent faibles. Ces premiers résultats sont donc à considérer comme des tendances à confirmer avec des cohortes plus importantes sur d'autres publics et sur d'autres TBA. Cette étude présente plusieurs limites : (1) le caractère optionnel du jeu dans le cursus qui réduit de fait l'impact du TBA, (2) l'impossibilité technique d'identifier précisément les moments de consultation des pages, diminuant le volume de données considérées (sessions de moins de 29 jours), (3) l'estimation de la lecture réelle des pages du TBA : il conviendrait de tester l'impact du temps de consultation des pages vues. Au-delà des analyses temporelles, certaines réponses des étudiants posent question (TBA jugé non utile bien que consulté plusieurs fois). Plusieurs interprétations sont possibles : (1) des étudiants ont une impression globale négative mais exploitent quelques indicateurs particuliers, (2) un effet d'oubli si le questionnaire arrive trop tard après l'usage. Une solution intégrée à chaque page du TBA et traçant le temps passé sur la page et le moment de consultation pourrait apporter de meilleurs résultats. Une autre piste serait de personnaliser le questionnaire

en informant l'utilisateur sur le nombre de consultations d'une page mais serait contraignante d'un point de vue RGPD.

Conclusion. Les TBA ne peuvent plus être pensés de manière statique comme le montrent les recherches actuelles autour des TBA adaptatifs [1]. Pour ouvrir des perspectives vers des modèles de TBA adaptatifs dans le temps, comme proposer une base de règles à coupler avec des adaptations selon les objectifs pédagogiques comme le proposaient Park et al.[4], nous avons essayé d'identifier les éventuels besoins d'adaptations des apprenants dans le temps. D'après les étudiants, certaines adaptations temporelles sont utiles comme la planification en début de semestre et l'évaluation en fin de semestre, et d'après leurs traces, nous pouvons observer des utilisations différentes dans le temps. Mais certains résultats sont contre-intuitifs, comme l'usage de la page d'évolution-révision qui semble être utilisée comme un outil de régulation de l'apprentissage, et demanderaient à être approfondie, par exemple via des focus-groupes. Si cette étude montre l'importance de croiser traces d'apprentissage et ressenti des utilisateurs, il reste possible d'affiner ces résultats en cherchant un éventuel lien entre l'évaluation du TBA, leur perception, les usages réels et les résultats universitaires. Nous pourrions aussi demander à l'apprenant de sélectionner les indicateurs utiles à plusieurs moments dans le semestre. Enfin, l'adaptation dans le temps du TBA pourrait aussi être explorée par rapport aux profils d'étudiants, selon leurs avancées ou leurs stratégies d'apprentissage.

Références

1. Ifenthaler, D. : Learning analytics for school and system management. In : OECD Digital Education Outlook 2021 : Pushing the Frontiers with Artificial Intelligence, Blockchain and Robots. OECD Publishing (2021)
2. Jivet, I., Scheffel, M., Drachsler, H., Specht, M. : Awareness Is Not Enough : Pitfalls of Learning Analytics Dashboards in the Educational Practice. In : Data Driven Approaches in Digital Education. pp. 82–96 (2017)
3. Oliver-Quelennec, K., Bouchet, F., Carron, T., Pinçon, C. : Analyzing the Impact of e-Caducée, a Serious Game in Pharmacy on Students' Professional Skills over Multiple Years. In : CSEDU 2021 - 13th International Conference on Computer Supported Education. vol. 1, pp. 331–338. SciTePress - Science and Technology Publications (2021)
4. Park, E., Ifenthaler, D., Clariana, R.B. : Adaptive or adapted to : Sequence and reflexive thematic analysis to understand learners' self-regulated learning in an adaptive learning analytics dashboard. British Journal of Educational Technology (2022)
5. Park, Y., Jo, I.H. : Factors that affect the success of learning analytics dashboards. Educational Technology Research and Development **67**(6), 1547–1571 (2019)
6. Roberts, L.D., Howell, J.A., Seaman, K. : Give Me a Customizable Dashboard : Personalized Learning Analytics Dashboards in Higher Education. Technology, Knowledge and Learning **22**(3), 317–333 (2017)
7. Verbert, K., Ochoa, X., De Croon, R., Dourado, R.A., De Laet, T. : Learning analytics dashboards : The past, the present and the future. Proceedings of the Tenth International Conference on Learning Analytics & Knowledge pp. 35–40 (2020)

Vers un outil de débriefing des simulations en réalité virtuelle

Kelly Minotti[1], Samir Otmane[1], Guillaume Loup[1]

[1]Université Paris-Saclay, Univ Evry, IBISC, 91020, Evry-Courcouronnes, France

Résumé.

La simulation et le débriefing sont deux phases indissociables pour les formations en réalité virtuelle. Avec la démocratisation de ces formations, il est essentiel d'identifier les approches pédagogiques pour le formateur et l'apprenant. Cependant, peu de travaux ont été dédiés aux spécificités des débriefings en environnement virtuel. Cet article propose une méthode immersive adaptable à différents types de formations en réalisant un état de l'art et une analyse des débriefings actuels en réalité virtuelle, suivie d'une méthode d'assistance au débriefing immersif.

Mots clés: Débriefing, Simulation, Réalité Virtuelle, Expérience, Formateur

Abstract.

Simulation and debriefing are two inseparable phases of virtual reality training. With the democratization of these pieces of exercise, it is essential to identify the pedagogical approaches for the trainer and the learner. However, only some studies have focused on the specificities of debriefings in a virtual environment. This article proposes an immersive method adaptable to different types of training by carrying out a state of the art and analysis of current debriefings in virtual reality, followed by an immersive debriefing assistance method.

Keywords: Debriefing, Simulation, Virtual Reality, Experience, Trainer

1 Introduction

La réalité virtuelle est devenue incontournable dans divers domaines tels que la santé et la formation militaire. Elle offre une expérience multisensorielle unique, un suivi et une analyse en temps réel, et peut renforcer la motivation et l'engagement de l'apprenant. Les simulateurs en réalité virtuelle sont utilisés pour confronter les apprenants à des situations exceptionnelles sans risques associés. Le principal objectif pédagogique est d'assurer l'acquisition de nouvelles compétences en mémorisant et interprétant les expériences grâce à des phases de simulation, de débriefing et de délibération. Le débriefing est considéré comme essentiel

dans l'apprentissage [1], mais peu de travaux ont été réalisés pour développer une méthode adaptée aux spécificités des simulateurs en réalité virtuelle. La recherche proposée vise à présenter une nouvelle méthode générique pour concevoir un environnement de débriefing, en analysant les différentes phases du débriefing et en identifiant les spécificités et limites des débriefings immersifs. Une première étude a été menée sur une approche similaire concernant les jeux sérieux [2].

2 Approche pédagogique du débriefing

Nous définissons l'apprenant comme participant à la simulation, le débriefeur comme formateur guidant le débriefing et le facilitateur comme débriefeur soulignant le contexte d'apprentissage. Le "débriefeur" est alors un "facilitateur de débriefing" pour atténuer toute perception d'autorité [3].

2.1 Définition du débriefing

L'analyse des définitions du débriefing a permis d'identifier ses caractéristiques et méthodes. À l'origine, le débriefing était une réunion technique de retour de mission militaire et un bilan psychologique pour S.Marshall [4]. Aujourd'hui, il est utilisé dans les secteurs militaires, aéronautiques, industriels, éducatifs et psychologiques. Sept définitions ont été retenues pour identifier la plus adéquate pour les simulateurs en réalité virtuelle.D'après ces définitions, l'objectif général du débriefing est de faciliter la réflexion guidée dans le cycle d'apprentissage expérientiel et post-expérience. Des définitions plus précises soulignent la présence d'une simulation, les notions de discussion, de réflexion et d'analyse. Enfin, une définition détaillée de la Haute Autorité de Santé met l'accent sur le rôle du débriefeur, absent dans les autres définitions : "Le débriefing est le temps **d'analyse** et de synthèse qui succède à la mise en situation simulée. C'est le temps majeur **d'apprentissage et de réflexion** de **la séance de simulation**. Il permet au formateur de **revenir sur le déroulement du scénario**, selon un processus structuré. L'idée est de dégager avec les apprenants les points correspondant aux objectifs fixés (éléments d'évaluation). Cette rétroaction (feedback) porte spécifiquement sur **l'analyse des performances** lors du déroulement du scénario et renseigne à la fois le formateur et l'apprenant. Le rôle du formateur est essentiel, car c'est lui qui va "faciliter" le débriefing et guider la réflexion des apprenants" [5].

2.2 Les caractéristiques d'un débriefing

Lederman [6] répertorie les éléments clés d'un débriefing efficace, tels que les participants, l'expérience de simulation, l'impact et le souvenir de la simulation, le support de débriefing, le temps de traitement et le facilitateur de débriefing.

Le moment choisi pour effectuer le débriefing aura une incidence sur la manière dont les participants vivront l'expérience. Bien qu'il soit possible de

débriefer en pleine simulation en la mettant en pause, il est préférable de le faire à la fin de la simulation, selon une étude montrant qu'une simulation sans interruption produisait un niveau plus élevé de réalisme clinique et émotionnel [7].

Le facilitateur de débriefing joue un rôle crucial en guidant les participants à travers une discussion ciblée, ce qui peut aider les apprenants à améliorer leurs compétences et leur apprentissage. Une étude portant sur la simulation de soins préopératoires a montré qu'un débriefing dirigé par un instructeur était plus efficace qu'un débriefing sans instructeur en termes de qualité du débriefing et de compétences de l'apprenant [8].

2.3 Phases pédagogiques du débriefing

Plusieurs méthodes de débriefing triphasiques sont couramment utilisées dans la littérature médicale, telles que le RAS (Reaction, Analyze and Summarize), le GAS (Gather, Analyze and Summarize), le modèle 3D et le Diamond [9]. Le RAS est la méthode la plus couramment utilisée et comprend trois phases. La première phase, appelée Réaction, permet aux participants d'exprimer leurs émotions juste après la séance d'apprentissage. La deuxième phase, appelée Analyse, permet d'enquêter et d'analyser les évènements importants du point de vue des participants, notamment leurs états d'esprit. Enfin, la dernière phase, appelée Résumé, consiste à examiner les compétences acquises. La méthode GAS, quant à elle, commence par la collecte d'informations pour établir un modèle mental partagé avant d'analyser les actions et de résumer les compétences acquises. Le modèle 3D est conçu pour répondre aux besoins individuels de l'expérience et de l'environnement, tandis que la méthode Diamond utilise une technique de description, d'analyse et d'application. Bien que ces méthodes soient associées à des termes différents, elles partagent des objectifs similaires : permettre l'expression des émotions, analyser les évènements et résumer les points à retenir.

Différentes techniques de débriefing existent, avec des nombres de phases variables. Par exemple, la technique Plus-Delta se limite à deux phases, tandis que la méthode PEARL comporte quatre phases. La méthode After Action Review, utilisée dans les domaines médical et militaire, comprend respectivement 7 et 10 étapes. Malgré ces différences, toutes ces techniques conservent les trois phases principales de réaction, d'analyse et de résumé pour exprimer les émotions, analyser les actions et faire un bilan des compétences acquises.

2.4 Limites des techniques de débriefing immersifs actuelles

Malgré l'évolution des simulateurs en réalité virtuelle, peu de travaux sont dédiés à l'usage de debriefing immersif, en particulier dans le domaine médical. Les méthodes de débriefing les plus courantes sont orales ou écrites, et l'emploi de la vidéo peut être utilisé pour la phase d'analyse. La vidéo 360° est la méthode la plus immersive à ce jour pour revoir les actions réalisées lors de la séance d'apprentissage. Une étude de Nicholson et al.[10] s'est concentrée sur son utilisation dans le débriefing et a interrogé ce que cet outil immersif peut apporter.

Les participants ont affirmé que la réalité virtuelle fournissait un environnement de débriefing plus immersif, dans lequel ils étaient plus engagés dans leurs performances et dont la vision des événements de simulation était meilleure.

3 Méthode d'assistance au débriefing immersif

Les simulations en réalité virtuelle ont des avantages pour les séances de débriefing. En effet, la possibilité de revoir la simulation en immersion permettrait à l'apprenant et au formateur de mieux se rappeler la séance d'apprentissage et de s'assurer qu'aucun événement clé n'a été oublié.

3.1 Choix de l'approche pédagogique

Le débriefing triphasique comme le RAS semble être l'approche pédagogique la plus utilisée pour les formations en réalité virtuelle. La première phase, où l'apprenant exprime ses émotions, nécessite un retour à la réalité pour une durée suffisante afin d'analyser l'expérience et créer un climat de confiance avec le débriefeur. Cependant, L'utilisation de la réalité virtuelle peut compliquer la dimension sociale des échanges entre le débriefeur et l'apprenant. La dernière phase, appelée phase de résumé, ne nécessite pas d'immersion et invite les apprenants à transposer les compétences acquises dans diverses situations. La phase qui nécessite l'utilisation de la réalité virtuelle est la phase d'analyse, qui permet à l'apprenant de revoir les événements pertinents de la simulation et de se remémorer plus facilement son ressenti. Ainsi, les recommandations de durée d'utilisation du casque de réalité virtuelle peuvent être respectées pour éviter une potentielle fatigue à l'issue de la session de débriefing.

L'outil proposé offre un débriefing immersif après une simulation, souvent utilisé dans la formation médicale ou militaire. Il permet aux apprenants d'observer leurs performances et émotions grâce à une rediffusion de la simulation dans la scène virtuelle, plutôt qu'en utilisant une vidéo 360°. Cette approche combine les avantages de la vidéo avec l'immersion offerte par la réalité virtuelle. L'objectif est de progresser vers l'assimilation et l'adaptation pour le transfert de l'apprentissage vers des situations futures. L'hypothèse est que le retour dans l'environnement virtuel améliorera le transfert de compétences. Pendant la rediffusion, l'apprenant pourra se déplacer librement dans la scène en utilisant un avatar de lui-même pour voir ses actions à la troisième personne, plutôt qu'à la première personne comme dans une vidéo 360°.

3.2 Limites de la réalité virtuelle

L'utilisation de la réalité virtuelle pour le débriefing peut être associée à des contraintes telles que la fatigue oculaire de l'apprenant qui peut être exacerbée lors d'une session de débriefing prolongée. Pour pallier cela, il est recommandé de ne pas utiliser le casque de réalité virtuelle pendant la phase de réaction pour permettre à l'apprenant de se reposer et de se concentrer. Les débriefings en

groupe peuvent être facilités par l'utilisation d'un environnement collaboratif avec des avatars pour chaque participant et débriefeur. La communication non verbale est également importante dans le débriefing, mais il est difficile de la percevoir lorsque les participants portent un casque de réalité virtuelle. Pour identifier les évènements clés, des indicateurs peuvent également être basés sur des mesures du rythme cardiaque, de l'analyse oculaire et de l'expression faciale pour fournir au débriefeur des informations sur les influences émotionnelles et le stress potentiel du sujet.

3.3 Architecture globale du système d'assistance au débriefing immersif

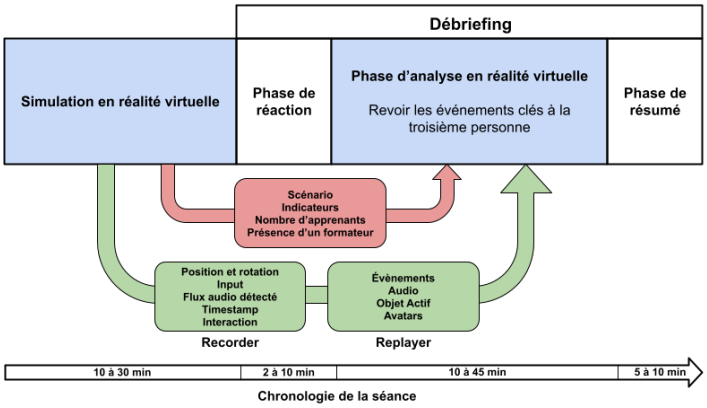


Fig. 1. Architecture globale du système d'assistance au débriefing immersif, avec chronologie de la séance [11]

La figure 1 montre l'architecture du système d'assistance au débriefing immersif. Le but est de récupérer des informations de la simulation (flèche rouge) pour concevoir une phase d'analyse immersive adaptée aux besoins de l'apprenant. Les éléments nécessaires comprennent le scénario, le nombre d'apprenants, la présence ou non d'un débriefeur, ainsi que les indicateurs. Les indicateurs seront basés sur le modèle de scénario, les données de la scène et des périphériques, et leur sélection sera confiée à une tâche dédiée.

La flèche verte dans la figure 1 correspond à notre système de replay. Pour enregistrer les éléments de la scène Unity, tels que la position et la rotation de l'utilisateur, les événements de la scène, l'audio ambiant et le flux audio capté par le casque, ainsi que les objets interactifs, l'algorithme Recorder est utilisé pour stocker toutes ces données dans un fichier de sortie. D'autre part, l'algorithme Replayer lit les données enregistrées et offre une fonction de navigation temporelle pour accéder à des moments spécifiques de la lecture. Pour une

optimisation de la mémoire, seuls les changements d'états seront pris en compte, ainsi que les objets statiques qui n'ont pas d'impact sur la simulation ne sont pas enregistrés.

4 Conclusion

Cet article vise à concevoir un outil de débriefing adapté aux simulations en réalité virtuelle en étudiant les définitions, caractéristiques et types de débriefings présents dans la littérature. La méthode la plus appropriée est un débriefing triphasique, dont seule la phase d'analyse nécessite un retour d'informations de la simulation.

Une étude est en cours auprès de formateurs dans le domaine militaire et médical. Un questionnaire sur l'approche pédagogique du débriefing (définition, caractéristiques, méthode) a été mis en place et distribué. Ceci permet d'apporter des compléments de terrain de notre approche. Leur retour permettra de concevoir un premier prototype qui sera évalué.

References

1. Savoldelli, G. L. *et al.* Value of Debriefing during Simulated Crisis Management. *Anesthesiology* **105,** 279–285 (2006).
2. Degand, J., Loup, G. & Didier, J.-Y. *Towards an Immersive Debriefing of Serious Games in Virtual Reality: A Framework Concept* in *Games and Learning Alliance* (Cham, 2021), 143–152.
3. Oriot, D. & Alinier, G. *La simulation en santé - Le débriefing clés en mains* (2019).
4. Crocq, L. Histoire du debriefing. *Pratiques Psychologiques* **10** (2004).
5. *Guide méthodologique : simulation en santé et gestion des risque* Feb. 2019.
6. Lederman, L. C. Debriefing: Toward a Systematic Assessment of Theory and Practice. *Simulation & Gaming* **23,** 145–160 (1992).
7. Van Heukelom, J. N., Begaz, T. & Treat, R. Comparison of Postsimulation Debriefing Versus In-Simulation Debriefing in Medical Simulation. *Simulation in Healthcare: The Journal of the Society for Simulation in Healthcare* **5,** 91–97 (2010).
8. Kim, S. S. & De Gagne, J. C. Instructor-led vs. peer-led debriefing in preoperative care simulation using standardized patients. *Nurse Education Today* **71,** 34–39 (Dec. 2018).
9. Abulebda, K., Auerbach, M. & Limaiem, F. Debriefing Techniques Utilized in Medical Simulation. *StatPearls* (Sept. 2022).
10. Nicholson, J., Gillespie, R., Bickerdike, S., Frith, G. & Hassan, T. OP13 Immersive video for simulation debriefing: 'record and review' in 360-degrees using a virtual reality headset. *BMJ Simulation and Technology Enhanced Learning* **5** (2019).
11. Savoldelli, G. & Boet, S. in *La simulation en santé De la théorie à la pratique* 313–328 (Springer Paris, 2013).

EIAH2023 : 11ème Conférence sur les Environnements Informatiques pour l'Apprentissage Humain
La conférence pluridisciplinaire francophone sur la conception et l'analyse des environnements numériques pour l'éducation et la formation
12-16 juin 2023 Brest (France)

Session de communications 5
Indicateurs et mesures

Mesurer le « flow » lors de l'utilisation d'une plateforme de jeux pédagogiques par des élèves du primaire

Aous Karoui[1], Lionel Alvarez[1,2], Thierry Geoffre[1], Quentin Brumeaud[1]

[1] Haute École Pédagogique de Fribourg, 1700 Fribourg, Suisse
prenom.nom@edufr.ch
[2] Université de Fribourg, 1700 Fribourg, Suisse
lionel.alvarez@unifr.ch

Résumé. Le concept de « flow » est utilisé pour décrire l'état mental d'être complètement immergé dans une activité difficile, mais réalisable. Bien qu'il soit étudié principalement dans le cadre des activités de loisirs, le « flow » suscite aujourd'hui de plus en plus l'intérêt des chercheurs dans le domaine des EIAH et de la *gamification*. Cet article présente les premiers résultats de mesures de « flow » lors de l'utilisation de la plateforme de jeux pédagogiques GamesHUB dans quatre classes d'école primaire (N = 76). Une version du questionnaire *E-GameFlow*, adaptée pour des enfants, a permis d'obtenir des notes élevées en termes de *concentration*, de *défi*, d'*autonomie* et d'*immersion* perçus par les élèves. De plus, les réflexions faites à travers les dessins et l'expression verbale des élèves montrent que la plateforme GamesHUB offre des opportunités pour atteindre le « flow » tout en apprenant. Malgré ces résultats prometteurs, des recherches supplémentaires sont nécessaires pour vérifier l'efficacité de l'outil, mettant ainsi en évidence des pistes potentielles de développements futurs.

Mots-clés : EIAH, apprentissage ludique, pédagogie du jeu, e-learning, évaluation du « flow »

Abstract. The concept of "flow", or the mental state of being fully immersed in a challenging but achievable activity, although studied primarily in the context of leisure activities, is now increasingly attracting the interest of researchers in the field of HIE and gamification. This article presents the first results of measuring "flow" following the use of the GamesHUB educational game platform in four elementary school classrooms (N = 76). A child-friendly version of the *E-GameFlow* questionnaire yielded high scores in terms of students' perceived concentration, challenge, autonomy, and immersion. In addition, reflections from the students' drawings and verbal expression show that the GamesHUB platform offers opportunities to achieve "flow" while learning. Despite these promising results, further research is needed to verify the effectiveness of the tool, highlighting potential avenues for future developments.

Keywords: TEL, game-based learning, e-learning, flow assessment.

1 Introduction

Les plateformes de jeux pédagogiques sont des outils numériques qui permettent aux élèves d'apprendre à travers des jeux interactifs et immersifs [1]. En jouant à ces types de jeux, les élèves peuvent explorer différents concepts et idées d'une manière engageante, les encourageant à développer des compétences en résolution de problèmes, leur pensée critique, leur créativité et la collaboration [2, 3]. Ainsi, l'apprentissage par le jeu peut aider à rendre l'apprentissage plus agréable et à augmenter la motivation et la réussite scolaire des élèves [4]. À ce titre, de plus en plus de chercheurs s'intéressent à la mesure du « flow » ressenti par les apprenants lors des sessions de jeux [5, 6].

En effet, le « flow » est un terme utilisé pour décrire l'état mental de l'expérience optimale, où les apprenants sont tellement immergés dans la tâche à accomplir qu'ils perdent la notion du temps et sont capables d'atteindre des performances optimales [7]. L'apprentissage par le jeu a le potentiel de faciliter l'atteinte du « flow » chez les apprenants et d'accroître ainsi l'efficacité de leur expérience d'apprentissage [8].

2 La plateforme GamesHUB

2.1 Les fondements de la plateforme

Débuté en 2020, le projet GamesHUB est né d'une collaboration entre la HEP de Fribourg et l'Ecole des Métiers de Fribourg (EMF). Le projet vise à concevoir et à développer une plateforme d'apprentissage en ligne pour les élèves d'école primaire, avec ou sans troubles du langage, qui soit autant que possible fondée sur des preuves et visant une ambition de Conception Universelle de l'Apprentissage (CUA) [9].

La plateforme GamesHub est ainsi portée par deux institutions pluridisciplinaires au sein de la HEP de Fribourg qui sont le Centre de Recherches pour l'Enseignement / Apprentissage par les TEchnologies (CRE /ATE) et l'unité de recherche Pluralité du Langage & des Médias (PL&M). Le CRE/ATE et PL&M disposent de deux équipes pluridisciplinaires (informatique, didactique du français, linguistique appliquée, psycholinguistique) qui contribuent au développement de la plateforme GamesHub. Plus récemment, les développements sur GamesHub ont particulièrement été orientés vers l'apprentissage ludifié des langues dans le cadre du projet PEAPL[1] (Plateforme Européenne pour l'Apprentissage des Langues), soutenu par le fonds européen Erasmus+[2].

2.2 Les fondements didactiques

GamesHUB donne accès à différents jeux et niveaux de jeu, tous liés à des objectifs pédagogiques. Dans une expérience de base, l'élève peut parcourir les jeux et les différents niveaux de difficulté, organisés par objectifs pédagogiques, sujets et niveaux scolaires. L'enseignant peut donner la tablette à l'élève pour qu'il joue librement afin

[1] http://peapl.eu/

[2] Erasmus+ KA201-302DDA9 (2020-2023, 449'923€). https://peapl.eu/

d'apprendre, ou bien lui assigner des listes de niveaux de jeu dans le cadre d'un parcours personnalisé.

En effet, quelques tests avec des étudiants futurs enseignants, en 2020, ont indiqué la nécessité de permettre une progression plus articulée de l'expérience d'apprentissage, proche du cadre d'une séquence didactique tel qu'elle pourrait être proposée dans une classe ordinaire. Par conséquent, des parcours d'enseignement/apprentissage peuvent désormais être développés par l'enseignant (via le tableau de bord de l'enseignant) en utilisant des jeux et des niveaux de jeu comme ressources pédagogiques choisies et organisées pour cibler un objectif d'apprentissage. Comme illustré par la figure 1, la sélection de différents niveaux de jeux d'apprentissage est rendue possible grâce à une série de filtres (élément du référentiel de compétences qui est visé, difficulté, tags) et la partie supérieure de l'écran permet de visualiser le parcours créé comme une chaîne de « briques », chaque brique étant un niveau de jeu. Un parcours peut également intégrer des liens vers des ressources en ligne ou des instructions pour poursuivre une « activité non connectée » (par exemple, travail en groupe, utilisation d'une ressource papier ou vérification auprès de l'enseignant).

Pour un parcours d'apprentissage créé, les remédiations peuvent être anticipées et ajoutées au parcours comme des briques optionnelles envoyées uniquement à l'apprenant en fonction du score obtenu à une ou plusieurs étapes précédentes (en mode évaluation). Cela permet d'ajouter des briques pour aider à retrouver de la compréhension ou cibler un prérequis selon un référentiel de compétences donné. Une fois le parcours entièrement créé, il peut être adressé à un ou plusieurs apprenants [10].

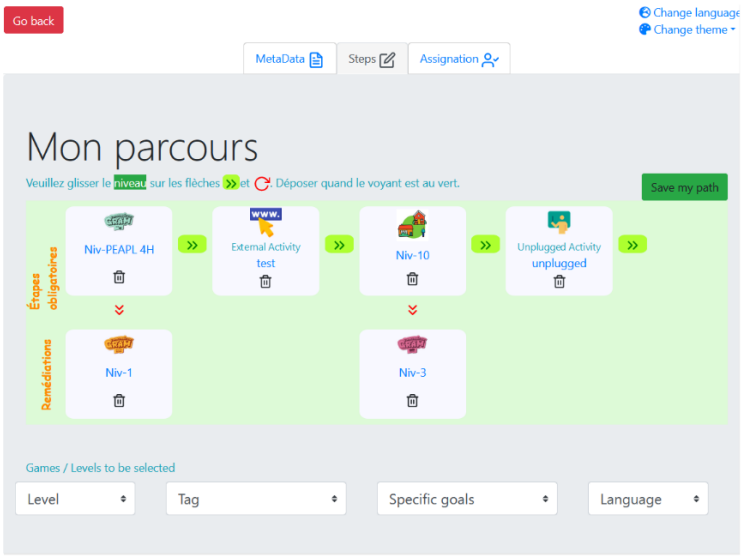


Fig. 1 Un exemple de parcours en cours de création sur GamesHUB, combinant des étapes obligatoires et des remédiations.

2.3 La conception de la plateforme

Afin de s'assurer que chaque apprenant, quelles que soient ses difficultés d'apprentissage, puisse bénéficier d'un engagement dans les parcours d'apprentissage ou les jeux de GamesHUB — dès lors qu'il peut interagir avec une tablette ou un ordinateur —, la plateforme a été conçue dans l'optique de la CUA. En conséquence, deux lignes directrices ont été choisies dès le début pour respecter cette ambition dans le développement : (1) la réduction des barrières, et (2) la diversité (des expériences d'apprentissage, des médiations, et des façons d'exprimer les compétences).

Le premier point, tel qu'illustré par la figure 2, consiste à fournir des outils d'assistance numérique qui sont des fonctionnalités additionnelles. Elles sont rendues disponibles à chaque situation d'apprentissage comme l'adaptation du contraste ou de la police de caractères, ainsi que la synthèse vocale, les blocs-notes ou le dictionnaire visuel [11]. En outre, la mise en œuvre d'un système d'apprentissage adaptatif qui recommande des jeux ou des niveaux de jeu est également en cours [12].

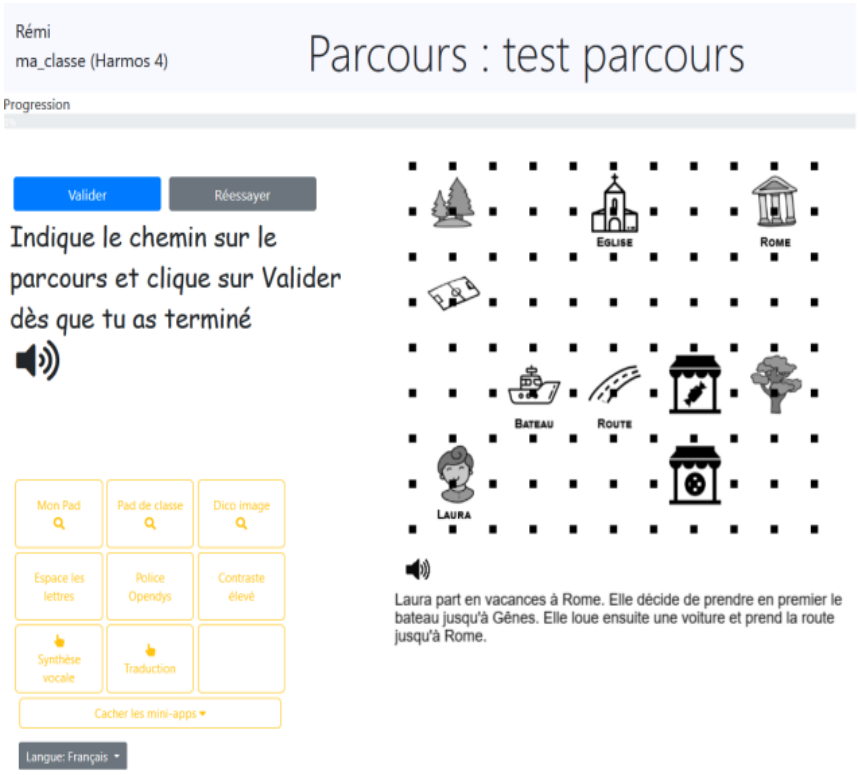


Fig.2. Une capture d'écran des fonctionnalités additionnelles disponibles sur GamesHUB

Le deuxième point — la gestion de la diversité — existe de différentes manières dans GamesHUB :

- Chaque jeu est développé avec quatre modes de jeu qui représentent quatre niveaux de taxonomie (explorer, former, évaluer, créer) [13].
- L'expérience de l'apprenant peut être un jeu autoguidé ou un parcours d'apprentissage personnalisé [14].
- Chaque jeu existe en plusieurs niveaux de difficulté.
- Dans cette pluralité de niveaux de difficulté, une variété d'antécédents culturels est explicitement choisie.
- L'alternance entre l'apprentissage en ligne et les activités non connectées est encouragée et rendue possible par des étapes simples au sein des parcours d'apprentissage créés par l'enseignant.
- L'intégration de ressources externes est soutenue pour permettre une plus grande diversité de médiations.

3 Tester le « flow » et l'expérience d'apprentissage

L'évaluation du « flow » dans l'enseignement, lors de l'utilisation d'une interface numérique, est relativement courante [15]. Le « flow » est généralement défini comme une expérience d'apprentissage optimale ou un « état mental psychologique d'une personne qui est immergée dans une activité avec une concentration énergique, un plaisir optimal, une implication totale et des intérêts intrinsèques, et qui est généralement concentrée, motivée, positive, énergique et alignée sur la tâche à accomplir » [16].

L'importance du « flow » pour l'apprentissage réside dans le fait que, lorsqu'une personne est en état de « flow », elle est motivée intrinsèquement et est plus encline à continuer à s'engager dans l'activité. De plus, les personnes en état de « flow » sont plus susceptibles d'apprendre de manière efficace, car elles sont plus concentrées et ont une plus grande capacité d'attention et de mémorisation.

3.1 Le questionnaire original : EgameFlow

Le questionnaire *EgameFlow* est un outil de mesure de la satisfaction des apprenants lors de l'utilisation de jeux pédagogiques [17]. Il s'agit d'une version modifiée du questionnaire GameFlow, qui a été développé pour mesurer l'expérience de « flow » dans les sessions de jeux vidéo [18].

D'un point de vue scientifique, le questionnaire présente de solides caractéristiques psychométriques, notamment une bonne validité de contenu et de construit. La *validité de contenu* mesure la capacité du questionnaire à évaluer tous les aspects pertinents de l'expérience de jeu, tandis que la *validité de construit* indique la capacité du questionnaire à mesurer ce qu'il est censé mesurer, c'est-à-dire la satisfaction de l'utilisateur. En outre, la fidélité du questionnaire a été évaluée et a montré une bonne cohérence interne des énoncés, suggérant que les énoncés mesurent tous la même chose et sont donc fiables. Le questionnaire a également montré une bonne sensibilité, capable de détecter les différences significatives entre les groupes de joueurs ayant des niveaux d'expérience différents ou ayant joué à des jeux différents.

3.2 Le questionnaire adapté : *EgameFlow* pour les enfants

L'*EgameFlow* original se compose de huit critères mesurés par une échelle de Likert à sept niveaux [17]. Cependant, dans notre contexte, nous avons supposé qu'il serait difficile pour des élèves de l'école primaire de fournir des réponses pertinentes à toutes les questions (42 au total). Nous avons donc créé une version adaptée de *l'EgameFlow*[3] pour ce public. Cette version inclut les quatre dimensions suivantes (parmi les huit initiales) : la concentration, le défi, l'autonomie et l'immersion. Comme les parcours d'apprentissage sont individuels, les critères d'interaction sociale ont été intentionnellement abandonnés. D'autres critères tels que l'amélioration des connaissances perçue et le feedback ont été abandonnés, car nous ne pouvions pas être sûrs de la pertinence des réponses données par les enfants. L'échelle de Likert de sept critères a été réduite à cinq, et les chiffres ont été remplacés par des smileys. Certaines des affirmations ont été reformulées afin de les rendre plus simples et plus compréhensibles par des enfants. Par exemple, l'affirmation « Je ne suis plus conscient de mon environnement pendant le jeu » pouvait être mal comprise par les enfants et prêter à confusion. Nous l'avons reformulée en « J'ai oublié le maitre/la maitresse et mes amis lorsque je faisais les exercices » pour rendre l'idée plus concrète.

À la fin, les enfants ont été invités à donner leur sentiment global sur l'utilisation de la plateforme, à travers un dessin et des mots-clés.

3.3 L'expérience d'apprentissage conçue

Trois parcours d'apprentissage ont été créés dans GamesHUB pour permettre ces premières sessions de test, selon les trois degrés scolaires impliqués (fin de 2e, 3e et 4e primaire). Nous avons utilisé des articulations progressives des niveaux de jeux les plus avancés actuellement disponibles sur la plateforme :

1. *Par ici ou par-là* (PCPL) qui implique un texte et un plan, donnant des instructions au lecteur pour se déplacer sur le plan (objectifs spécifiques : lecture et compréhension des verbes de mouvement et des connecteurs de lieu) ;
2. *L'Orthodyssée des Gram* (Gram) qui implique d'utiliser les indices morphosyntaxiques pour construire des phrases grammaticales.

Comme *Par ici ou par-là* est un jeu intégré à GamesHUB, nous avons souhaité évaluer comment les enfants se comporteraient avec le jeu, mais aussi avec le cadre habituel de la plateforme. À l'inverse, L'*Orthodyssée des Gram* [19] est un jeu qui a été développé précédemment, de manière indépendante, puis intégré à la plateforme ; nous voulions donc vérifier comment les élèves se débrouilleraient avec un parcours alternant ces deux jeux aux mises en page et au design différents. Des remédiations ont été ajoutées pour vérifier qu'elles étaient bien envoyées au moment voulu.

3.4 Les classes et la collecte des données

Les écoles où GamesHUB a été mise en œuvre pour décrire le « flow » expérimenté se trouvent en Romandie, la région francophone de la Suisse. Généralement, les classes

[3] https://blog.hepfr.ch/create/gameshub/k12-EgameFlow/ (accéder à la version adaptée du questionnaire)

sont composées d'une vingtaine d'élèves, et la pratique de l'enseignant est guidée par un programme appelé Plan d'Études Romand[4] qui liste toutes les compétences que les élèves doivent maitriser par degré scolaire. L'équipement informatique est très diversifié et inégal d'une école à l'autre, car les ordinateurs ou les tablettes sont payés par la ville ou le village et non par l'État. Quatre classes d'élèves de 8 à 10 ans (17 élèves dans une classe de 4H[5], 22 élèves dans une classe de 5H[6], et 37 dans deux classes de 6H[7]) avec suffisamment de tablettes pour chaque élève (N=76, 37 filles et 39 garçons) étaient disponibles pour la mise en œuvre et le test de GamesHUB.

Tout d'abord, une courte présentation a été faite sur les raisons de la présence du chercheur dans la classe. Ensuite, les élèves ont eu le temps de poser des questions sur GamesHUB et les fonctionnalités incluses. Ensuite, ils ont pu accéder à la plateforme web sur les tablettes. Les élèves disposaient d'un parcours d'apprentissage spécifique adapté à leur programme (un pour les élèves de la classe 4H, un pour ceux de la classe 5H et un pour les classes de classe 6H) pour apprendre avec GamesHUB, avant de devoir remplir le questionnaire *EgameFlow* adapté, imprimé sur papier. Pour s'assurer que les élèves avaient bien compris les questions, qui pouvaient être un défi, notamment pour les élèves de la classe 4H, nous leur avons d'abord expliqué le principe lors d'une courte séance plénière. Une assistance était ensuite possible pour les enfants qui pouvaient rencontrer quelques difficultés avec le français, comme les élèves allophones.

4 Le « flow » rapporté

4.1 Résultats

Cette section présente les données dans les tableaux 1 et 2, et la figure 3. Les statistiques descriptives ont été réalisées avec JASP 0.16.3[8], ainsi que la corrélation entre les dimensions retenues de *l'EgameFlow*.

Table 1. Statistiques descriptives du « flow » documenté avec l'*EgameFlow adapté*.

Variables	N	Moyenne	Std. Dev.	Min.	Max.
A. Concentration	76	4.40	0.68	2	5
B. Défi	76	4.33	0.62	1.5	5
C. Autonomie	76	4.47	0.79	2	5
D. Immersion	76	4.08	0.84	1	5

4 www.plandetudes.ch
5 2e primaire
6 3e primaire
7 4e primaire
8 https://jasp-stats.org/download/

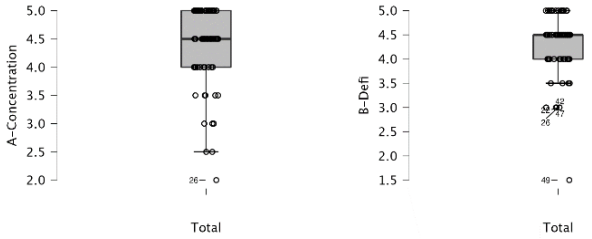


Fig. 3. Boxplots présentant le « flow » autodéclaré par les élèves, selon les 4 dimensions de l'EgameFlow adapté, après une session d'apprentissage de 2 fois 30′ sur GamesHUB.

Table 2. Corrélations de Pearson entre chaque dimension d'*EgameFlow*.

Variable		A-Concentra.	B- Défi	C- Autonomie	D-Immersion
A- Concentra.	r	-			
	p-value	-			
B- Défi	r	0.546***	-		
	p-value	< .001	-		
C-Autonomie	r	0.230*	0.508***	-	
	p-value	0.023	< .001	-	
D-Immersion	r	0.373***	0.551***	0.374***	-
	p-value	< .001	<. 001	< .001	-

Note. Tous les tests sont unilatéraux, pour une corrélation positive (* p<.05, ** p<.01, *** p<.001

4.2 Discussions

Les quatre dimensions évaluées par notre questionnaire *EgameFlow* adapté pour enfants sont élevées, comme le montre le tableau 1, avec une moyenne de 4,4 (SD=0,68) pour la concentration, de 4,33 (SD=0,62) pour le défi, de 4,47 (SD=0,79) pour l'autonomie et de 4,08 (SD=0,84) pour l'immersion. Cela pourrait indiquer un « flow » relativement intense ressenti par les élèves de 8 à 10 ans pendant qu'ils travaillent sur GamesHUB. Néanmoins, cela peut être dû à plusieurs facteurs. En effet, le fait de jouer sur des iPads et de faire l'expérience d'un nouvel environnement d'apprentissage rend plus difficile la séparation des facteurs de motivation liés au matériel et au contexte de ceux liés à la plateforme et aux jeux d'apprentissage eux-mêmes. Cependant, nous pensons que la conception actuelle de la plateforme GamesHUB ainsi que la conception des jeux *Orthodyssée des Gram* et *PCPL* ont contribué à ce sentiment général de satisfaction. D'autre part, ces résultats pourraient également indiquer un biais de désirabilité sociale [20], les réponses des élèves pouvant être influencées par la présence du chercheur lors des sessions de jeux pour collecter les données.

La corrélation entre chaque dimension est également élevée, systématiquement significative ($p < 0,5$ pour la corrélation entre la concentration et l'autonomie, $p < 0,001$ pour toutes les autres corrélations) comme le montre le tableau 2. Cela pourrait également indiquer que les dimensions *EgameFlow* sont généralement corrélées, comme l'ont déclaré Fu et al. [17].

Finalement, l'hypothèse du biais de désirabilité sociale est plutôt à écarter, car nous avons demandé aux élèves de conclure délibérément leurs réponses au questionnaire par leur sentiment général, à la suite de cette expérience. Il leur a été demandé de soumettre leur expérience d'apprentissage en dessinant ce dont ils se souvenaient des jeux qu'ils préféraient et en écrivant des mots courts qui reflétaient leur état d'esprit à ce moment.

Les dessins que les élèves de 8 à 10 ans ont réalisés après avoir joué et appris avec GamesHUB (**Fig.4**) représentaient parfois le plaisir d'utiliser une tablette, mais montraient généralement qu'ils avaient apprécié les jeux *PCPL* et *L'Orthodyssée des Gram* intégrés dans les parcours d'apprentissage. Le fait que l'apprentissage soit basé sur le jeu semble bien être un facteur de motivation.

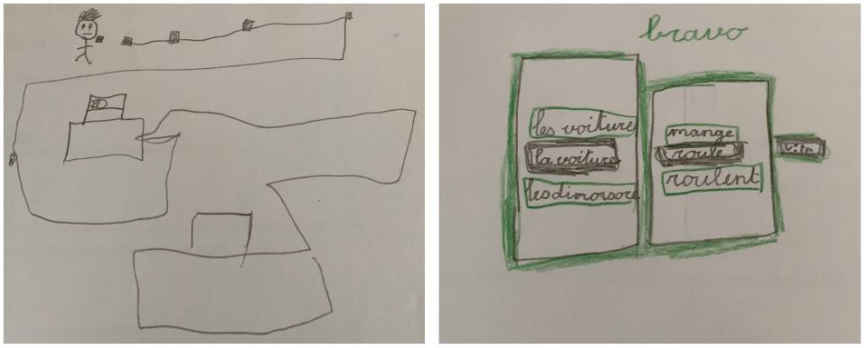


Fig. 4. A- Une représentation du jeu PCPL B- Une représentation de L'Orthodyssée des Gram

Les mots-clés (Fig. 5) montrent que les élèves ont vécu des expériences d'apprentissage et de plaisir grâce à la tablette et aux jeux conçus (PCPL et *Gram*). Certains ont indiqué l'intérêt de pouvoir refaire des tâches simples dans une progression. Les enseignants présents lors des expériences ont également déclaré que les élèves étaient plus motivés car ils étaient certainement attirés par cette façon d'apprendre.

Fig. 5. Nuage de mots présentant les expériences rapportées par les étudiants avec des mots-clés après avoir joué et appris avec GamesHUB.

5 Conclusion et perspectives

GamesHUB est une plateforme web d'apprentissage par le jeu, développée dans une optique de Conception Universelle d'Apprentissage (CUA). Une équipe interdisciplinaire en assure la conception et le développement. Les premiers tests en situation réelle (4 classes, 76 élèves) ont été réalisés pour évaluer la validité écologique grâce à notre version du questionnaire *EgameFlow* adaptée pour enfants. Les résultats font état d'un « flow » très positif exprimé par les élèves, dans chacune des quatre dimensions analysées.

Nous accueillons ces premiers résultats avec prudence en raison du possible biais de désirabilité sociale. Cependant, les retours directs, en observant l'activité des élèves et l'expérience rapportée à travers les dessins et les mots-clés, sont également positifs et encourageants pour poursuivre un processus itératif de développement et de tests.

D'autres études sont nécessaires pour documenter clairement l'expérience d'apprentissage des élèves, l'expérience des enseignants, et si l'ambition d'une CUA est explicitement mise en œuvre. Une nouvelle session de test, prévue dans le cadre du projet PEAPL, a été menée en mars 2023, avec entretiens avec les enseignants, avant une mise en œuvre plus large et plus ambitieuse incluant l'apprentissage adaptatif.

Remerciements : Nos plus vifs remerciements s'adressent aux enseignantes et enseignants des écoles fribourgeoises pour leur implication dans les expérimentations et leur contribution à ce travail de recherche.

Références

1. Li, Y., Xu, Z., Hao, Y., Xiao, P., Liu, J.: Psychosocial Impacts of Mobile Game on K12 Students and Trend Exploration for Future Educational Mobile Games. *Frontiers in Education*, *7* (2022). https://doi.org/10.3389/feduc.2022.843090
2. Dickey, M. D.: K-12 teachers encounter digital games: a qualitative investigation of teachers' perceptions of the potential of digital games for K-12 education. *Interactive Learning Environments*, 23, 485 – 495 (2015). https://doi.org/10.1080/10494820.2013.788036

3. Henríquez, V., Scheihing, E., Silva, M.: Incorporating Blended Learning Processes in K12 Mathematics Education Through BA-Khan Platform. *European Conference on Technology Enhanced Learning* (2018). https://doi.org/10.1007/978-3-319-98572-5_26
4. Can, T., Kucuk, S., Simsek, I.: Examining K-12 Students' Preferences and Attitudes on Mobile Learning. *International Journal of Technology in Teaching and Learning* (2019).
5. Wilfried A., Jantina H., Sanne A., Geert ten D.: The concept of flow in collaborative game-based learning. *Computers in Human Behavior*, *27*(3), 1185–1194 (2011). https://doi.org/10.1016/j.chb.2010.12.013
6. Lin, Y., Hou, H.T.: *The Design of an Ecosystem-Education Board Game Integrating Role-Play and Peer-Learning Mechanism and Its Evaluation of Learning Effectiveness and Flow* (2016). https://papers.iafor.org/wp-content/uploads/papers/ace2016/ACE2016_32767.pdf
7. Pavlas, D.: A Model of Flow and Play in Game-based Learning the Impact of Game Characteristics, Player Traits, And Player States. *Electronic Theses and Dissertations*, 2004-2019. (2010). https://stars.library.ucf.edu/etd/1657
8. Rachmatullah, A., Reichsman, F., Lord, T., Dorsey, C., Mott, B.W., Lester, J.C., Wiebe, E.N.: Modeling Secondary Students' Genetics Learning in a Game-Based Environment: Integrating the Expectancy-Value Theory of Achievement Motivation and Flow Theory. *Journal of Science Education and Technology*, 30, 511–528 (2021). https://doi.org/10.1007/s10956-020-09896-8
9. Meyer, A., Rose, D.H., Gordon, D.: Universal Design for Learning. Theory and practice. CAST incorporated, Wakefield, MA (2014). https://publishing.cast.org/catalog/books-products/universal-design-for-learning-meyer-rose-gordon
10. Karoui, A., Alvarez, L., Geoffre, T., Dherbey Chapuis, N., Rodi, M., Ramalho, M.: Adaptive Pathways within the European Platform for Personalized Language Learning PEAPL. In: Adjunct Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization. pp. 90–94. ACM, Utrecht, Netherlands (2021). https://doi.org/10.1145/3450614.3464480
11. Rodi, M., Geoffre, T.: Des séquences d'acquisition/apprentissage métagraphiques au sein d'un jeu en ligne: l'Orthodyssée des Gram. *Bulletin suisse de linguistique appliquée*, *1*, 77-102, (2021). https://www.vals-asla.ch/fileadmin/user_upload/Journal/Special_2021_vol_1_def_A4.pdf
12. 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. EC-TEL. (2022). https://doi.org/10.1007/978-3-031-16290-9_35
13. Anderson, L.W., Krathwohl, D.R.: A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives. Longman (2001).
14. Karoui, A., Alvarez, L., Geoffre, T., Dherbey Chapuis, N., Rodi, M., Ramalho, M.: Adaptive Pathways within the European Platform for Personalized Language Learning PEAPL. UMAP 2021 - Adjunct Publication of the 29th ACM Conference on User Modeling, Adaptation and Personalization. 90–94 (2021). https://doi.org/10.1145/3450614.3464480
15. Alvarez, L., Carrupt, R., Audrin, C., Gay, P.: Self-Reported Flow in Online Learning Environments for Teacher Education: A Quasi-Experimental Study Using a Counterbalanced Design. *Education Sciences. 12*, 351 (2022). https://doi.org/10.3390/educsci12050351
16. Bonaiuto, M., Mao, Y., Roberts, S., Psalti, A., Ariccio, S., Cancellieri, U.G., Csikszentmihalyi, M.: Optimal experience and personal growth: Flow and the

consolidation of place identity. *Frontiers in Psychology. 7*, 1–12 (2016). https://doi.org/10.3389/fpsyg.2016.01654

17. Fu, F.L., Su, R.C., Yu, S.C.: EGameFlow: A scale to measure learners' enjoyment of e-learning games. *Computers and Education. 52*, 101–112 (2009). https://doi.org/10.1016/j.compedu.2008.07.004
18. Sweetser, P., Wyeth, P. (2005). GameFlow: a model for evaluating player enjoyment in games. *Computers in Entertainment (CIE)*, *3*(3), 3-3. https://doi.org/10.1145/1077246.1077253
19. Geoffre, T., Hofer, D., Cochard, B. (2021). *L'Orthodyssée des Gram*. HEP|PH FR. www.lafamillegram.ch/#/
20. Krumpal, I.: Determinants of social desirability bias in sensitive surveys: a literature review. Qual Quant. 47, 2025–2047 (2013). https://doi.org/10.1007/s11135-011- 9640-9

De la modélisation des indicateurs pour l'apprentissage à la capitalisation par les utilisateurs

Albane Gril[1,2] Valérie Renault[2], Madeth May[1], Sébastien George[1]

[1] Le Mans Université, LIUM, F-72000 Le Mans, France
[2] Le Mans Université, CREN, F-72000 Le Mans, France
{albane.gril,valerie.renault,madeth.may,sebastien.george}@univ-lemans.fr

Abstract. Les indicateurs pour l'apprentissage, comme données apportant du sens dans le cadre de la formation, sont aujourd'hui utilisés dans une majorité des EIAH. Or, entre les importantes ressources nécessaires à leurs conceptions et les difficultés d'adoption par les utilisateurs, il peut être difficile de conserver leurs pertinences dans le temps. C'est à cette question que s'intéressent nos travaux. La capitalisation des indicateurs est ici proposée comme résultant de l'interaction des utilisateurs avec les indicateurs tout au long du cycle de vie de ces derniers. Cela nécessite tout d'abord de modéliser les indicateurs comme des entités capables d'une évolution selon l'usage qui en est fait. De plus, cette approche a ses propres défis d'adoption, car les utilisateurs, particulièrement les non-experts, ne sont pas forcément à l'aise avec ces notions abstraites. C'est pourquoi nous avons apporté un découpage par scénarios d'utilisations résultant d'interactions guidées des utilisateurs avec des indicateurs. Cette proposition est implémentée dans un prototype informatisé afin de permettre une expérimentation concrète par les utilisateurs.

Keywords: learning analytics · indicateur · capitalisation.

Abstract. Indicators for learning, as data providing meaning in the context of learning, are currently used in a majority of TEL systems. However, between the important resources required for their design and the difficulties of adoption by users, it can be difficult to maintain their relevance over time. This is the issue that our work addresses. The capitalization of indicators is presented here as a result of the interaction of users with indicators throughout their life cycle. This requires that indicators be modeled as entities capable of evolving according to their use over time. Moreover, this approach has its own adoption challenges, as users, especially non-experts, are not necessarily comfortable with these abstract notions. This is why we have provided a breakdown by usage scenarios resulting from guided interactions of users and indicators. This proposal is implemented in a computerized prototype in order to allow a concrete artifact experimentation by the users.

Keywords: learning analytics · data indicator · capitalization.

1 Introduction

L'utilisation d'indicateurs d'apprentissage dans le cadre de tableaux de bord intégrés à des Environnements Informatiques pour l'Apprentissage Humain (EIAH) est une pratique courante : de nombreuses études [1, 19] ont montré leur impact positif sur l'apprentissage. Dans le cadre du projet écri+ pour l'amélioration des compétences du français écrit des étudiants à l'université, la plateforme de formation et de certification dédiée génère de nombreuses traces. Elles sont actuellement peu exploitées, mais ont le potentiel de proposer aux étudiants comme aux enseignants des indicateurs pertinents sur les apprentissages.
Notons que la notion d'"indicateur d'apprentissage", tel que nous l'entendons, est une terminologie simplifiée des "indicateurs dans le contexte de l'apprentissage". Dans ce contexte, les indicateurs regroupent aussi bien des informations sur l'avancement dans un cours telles que la qualité de rédaction d'un étudiant, que le comportement de l'étudiant dans un environnement informatique tel que la fréquence de ses connexions. L'intégration de nouveaux indicateurs à un environnement d'apprentissage est complexe et coûteuse [3].
Nous considérons, dans cet article, l'intégralité du cycle de vie de l'indicateur, de sa conception impliquant les utilisateurs à différents niveaux, à son évolution dans le temps. Nous nous interrogeons sur comment permettre aux indicateurs d'évoluer dans le temps au gré des usages des utilisateurs et de la multiplicité des contextes ? L'objectif de notre travail est de proposer une approche dite de capitalisation des indicateurs d'apprentissage, c'est-à-dire mettant en avant le processus d'amélioration continue des indicateurs en favorisant les interactions des utilisateurs avec ces derniers.
Pour cela, plusieurs étapes clés ont déjà été effectuées. Tout d'abord, un recueil de besoins sur les interactions, que peuvent et veulent avoir nos utilisateurs cibles, a été réalisé. Cela a conduit à la constitution d'un corpus initial d'indicateurs, puis à la proposition d'une modélisation informatique des entités mises en jeu lors de la capitalisation, détaillée en section 3. De plus, nous avons pu préciser les interactions entre les entités et les utilisateurs pour permettre la capitalisation. Ces contributions théoriques ont ensuite été implémentées dans un prototype, sous forme d'application web, présenté en section 4 afin d'avoir des retours utilisateur sur le processus complet proposé, dont la première itération est discutée en section 5. L'expérimentation de ce prototype permet la validation de notre approche de capitalisation, mais aussi d'avancer avec les participants à la construction d'un outil pour instrumenter la capitalisation. L'analyse de ces retours va permettre une seconde phase de prototypage, pour affiner les interactions proposées et l'accompagnement des différents utilisateurs-acteurs. Nous allons exposer les enjeux, notre démarche et nos propositions initiales sur la capitalisation d'indicateurs dans la première section.

2 La capitalisation des indicateurs d'apprentissage

2.1 Défis techniques

La capitalisation répond tout d'abord à des défis techniques inhérents aux indicateurs d'apprentissage. Ils sont définis comme des observables signifiants [11, 13] sur le plan pédagogique, et témoignant de la qualité de l'interaction, de l'activité et de l'apprentissage dans un EIAH. Ils sont calculés ou établis à partir de traces récoltées dans les EIAH.
L'utilité des indicateurs, ainsi que leur grande diversité, est avérée dans le cadre des apprentissages [1, 19]. En effet, les données utilisées, leurs analyses et leurs paramètres varient en fonction du contexte et de l'objectif de l'indicateur, influençant sa conception. De même, la visualisation a un impact important sur la compréhension et l'usage d'un indicateur [20]. Ce processus de conception est très peu généralisable et reproductible, notamment à cause de la diversité des contextes d'apprentissage. Un indicateur n'est à priori pas conçu pour être diffusé et modifié par les utilisateurs. Pourtant, les besoins des utilisateurs peuvent changer au cours du temps et en fonction de leurs pratiques. Nous proposons donc que les indicateurs soient flexibles pour suivre ces évolutions.
Le processus de capitalisation des indicateurs prenant ainsi en compte leur possible évolution afin qu'ils conservent leur pertinence selon le contexte.

2.2 Défis humains

La conception des indicateurs d'apprentissage nécessite la collaboration de plusieurs expertises pour un résultat pertinent et fonctionnel [12]. En particulier, des notions d'analyse de données sont nécessaires, mais aussi d'expertise pédagogique.
Consulter les utilisateurs finaux des indicateurs, enseignants ou étudiants, pour un recueil de besoin initial est largement utilisé [7, 10, 15]. Or cette consultation est coûteuse en temps et en ressources, et donc souvent limitée à un cycle de conception initiale. Le côté ponctuel de certaines expérimentations laisse peu de temps aux utilisateurs pour s'approprier l'environnement et les indicateurs dans leurs pratiques d'enseignement et d'apprentissage. De même, l'absence d'expertise technique des utilisateurs est un défi lors d'un recueil de besoin [17].
Une fois un indicateur conçu au terme d'itérations de co-construction entre experts et utilisateurs, son adoption par la communauté n'est ni garantie ni nécessairement concordante par rapport à la diversité des pratiques réelles [2]. Les questions de dérive d'utilisation, d'accessibilité des données et autres préoccupations éthiques freinent également les utilisateurs dans leurs adoptions des indicateurs [16].
La diversité des expertises et des expériences individuelles complique la conception d'indicateurs adaptés aux besoins des individus. L'utilisateur doit donc pouvoir apporter à tout moment son point de vue sur l'indicateur, afin d'en assurer la pertinence pour son usage.

2.3 Définition d'une capitalisation par les utilisateurs

Sur l'intégralité du cycle de vie de l'indicateur, l'implication des utilisateurs finaux est un élément clé afin d'obtenir un indicateur pertinent et de faciliter son adoption. Cette implication des utilisateurs a une dimension temporelle importante dans l'évolution potentielle de l'usage d'un indicateur. Nous proposons donc de considérer la capitalisation comme une approche résultant de quatre actions de ces utilisateurs sur les indicateurs. Ces actions peuvent être associées à une scénarisation des interactions.
Cette proposition a été initiée par un état de l'art des approches de conception d'indicateurs dans la littérature [8] afin de prendre connaissance de l'existant en matière de solution de conception d'indicateurs. Cela a mis en lumière un manque de soutien et de solutions techniques adaptées aux utilisateurs non experts dans des actions de capitalisation d'indicateurs d'apprentissage.
La première action est l'*appropriation*, un concept souvent utilisé en sciences humaines et dans le domaine de l'IHM (Interaction Humain-Machine) [18]. Elle permet aux utilisateurs d'appréhender et de personnaliser les indicateurs sans nécessiter l'intervention d'experts. L'action de *réutilisation* implique que les indicateurs fonctionnent de la même manière dans un contexte similaire et peuvent ainsi être utilisés à nouveau sans altération. L'action d'*adaptation* nécessite la modification des indicateurs afin de les modifier pour une meilleure adéquation à un autre contexte ou de répondre à des besoins similaires. Enfin, l'action de *partage* permet aux utilisateurs d'accéder aux indicateurs capitalisés et de les diffuser au sein d'une communauté de pratique avec des retours constructifs.
Les quatre actions sont liées à l'usage qui est fait de l'indicateur. Nous proposons ainsi dans la partie suivante un modèle prenant en compte cet usage.

3 Modélisation des indicateurs pour leurs interactions avec les utilisateurs

Notre problématique est de rendre capitalisé et capitalisable tout indicateur d'apprentissage. Il est donc essentiel que sa conception permette à la fois une généralisation pour favoriser son utilisation et sa réutilisation, mais aussi une spécification permettant son appropriation.
De nombreux travaux portant sur la définition, puis la modélisation des indicateurs ont été réalisés au cours des dernières années, notamment avec l'augmentation conséquente de leur utilisation. Parmi tous ces travaux, nous relevons particulièrement ceux de Dimitrakopoulou [6] définissant un indicateur comme une variable, au sens mathématique, à laquelle est attribuée une série de caractéristiques, incluant des informations de contexte et d'utilisation. Diagne [5] propose une carte d'identité qui inclut le contexte de l'indicateur pour en définir le besoin auquel il répond, dans le but de faciliter la réutilisation. UTL (*Usage Tracking Language*) proposé par Iksal [11] est un méta-langage XML complet qui permet la modélisation des entités composant les indicateurs et facilite la phase d'expression des besoins en l'intégrant au modèle. Enfin, Lebis *et al.* [14] élaborent un processus narratif de conception du processus d'analyse de traces,

afin de faciliter les expressions par les concepteurs et ainsi leur capitalisation.
Tous ces travaux soulignent l'importance d'une définition claire de l'indicateur pour en permettre la capitalisation. Or l'indicateur n'y est pas considéré comme un objet défini et capitalisé directement par des utilisateurs. Notre intention est de commencer par modéliser ce qu'est un indicateur capitalisé. De quoi est-il constitué, quelles sont ses évolutions dans le temps, au cours des interactions avec les utilisateurs ? Nous allons maintenant détailler les deux entités qui sont proposées afin de mettre en avant la place centrale de l'utilisateur dans la définition et l'évolution de l'indicateur (voir Fig. 1) : l'objet *Indicateur* et l'objet *Usage*.

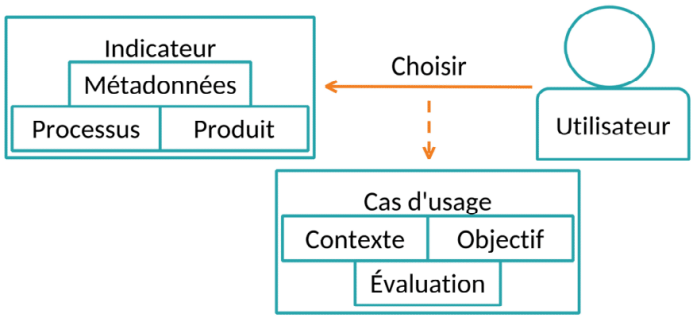


Fig. 1. Proposition de modélisation d'un indicateur capitalisé.

3.1 L'objet Indicateur

Cette modélisation se concentre sur les indicateurs et leurs capacités à évoluer au cours des utilisations. Le premier objet est donc l'Indicateur. Il est d'abord composé d'une partie de métadonnées telles qu'un titre, identité du concepteur, date de conception, ou encore une rapide description. Cette partie descriptive permet de garder des informations concernant la conception (où, quand, qui). Ces traces sont importantes pour la gestion des entités et des permissions lors de modifications d'indicateurs.
Il est ensuite constitué d'une partie nommée *Processus* qui a pour but de décrire le processus de conception de l'indicateur (comment), son analyse et les entités qui le constituent. La dernière partie est le *Produit*, qui est le résultat visible de l'indicateur (quoi). Il est conçu à partir de l'analyse décrite dans le processus. Les paramètres utilisés pour le calcul, et éventuellement des paramètres pour une visualisation de l'indicateur, peuvent aussi y être ajoutés.
Ces deux dernières parties sont complémentaires et essentielles à la définition d'un indicateur capitalisé. Elles correspondent à ce qui est conservé et considéré comme un objet capitalisé. Un indicateur doit donc être calculé au moins une fois pour être capitalisé. Par calculer, nous entendons ici que le processus d'analyse est utilisé sur des traces réelles afin d'obtenir un résultat exploitable et visualisable. Un indicateur calculé est ensuite celui choisi par un nouvel utilisateur pour son usage, et non un processus d'analyse de données seul. Cette

étape de calcul est indépendante de l'utilisateur et nécessite un moteur de calcul d'indicateurs. Ce moteur n'étant pas l'objet du présent article.
La capitalisation d'indicateurs suppose l'existence d'une base d'indicateurs capitalisés initiale. Cette base sert ainsi de point de départ aux utilisateurs. La figure 2 illustre l'utilisation du modèle. L'*Indicateur* y est capitalisé, dans le sens calculable et calculé. Or l'évolution des besoins auquel il répond et les contextes dans lesquels il s'inscrit n'est possible que si l'usage de l'indicateur peut être démultiplié pour un même indicateur capitalisé.

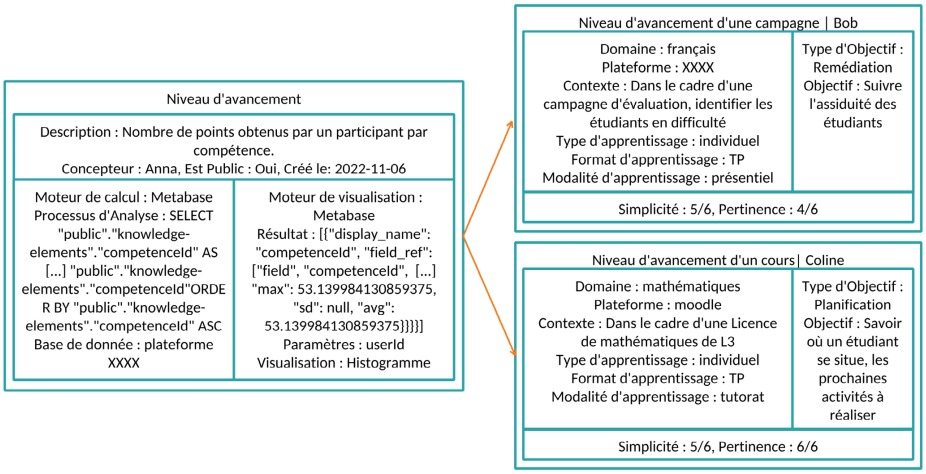


Fig. 2. Exemple d'utilisation de la modélisation d'un indicateur capitalisé : niveau d'avancement et deux instances d'Usages de cet indicateur.

3.2 L'objet Usage

L'ajout de l'objet *Usage* donne la possibilité d'avoir une infinité d'*Usage* pour un même *Indicateur*, favorise la diversification des interprétations et prend en compte la variabilité des utilisateurs essentielle à la capitalisation. Cette entité fait ainsi le lien entre l'utilisateur et l'indicateur. Choisir un *Indicateur* crée l'entité *Usage*, unique à un utilisateur et indépendant de l'*Usage* d'un autre avec le même *Indicateur*. Il est ainsi possible pour l'utilisateur d'interagir avec l'*Indicateur* sans l'altérer et en ajoutant des informations qui sont propres à son usage personnel.
L'*Usage* de l'indicateur d'apprentissage possède un *contexte* pédagogique déterminé par l'utilisateur. Il donne un cadre précis à l'indicateur, permettant son appropriation dans l'usage nouvellement créé. L'accumulation des contextes proposés guide les prochains utilisateurs dans le choix d'un indicateur pertinent. L'*Usage* est également lié à un *objectif* tel que défini par Dabbedi *et al.* [4], c'est-à-dire le type de décision prise à l'aide de l'information fournie par l'indicateur. Ainsi un même *Indicateur* peut avoir une grande diversité d'interprétation, d'objectif, sans nécessiter une duplication de l'entité. Enfin, l'entité recueille l'*évaluation* de l'usage de l'*Indicateur* lié. Cette évaluation, en partie renseignée

directement par l'utilisateur, est une information précieuse de la capitalisation. Elle permet une valorisation de l'expérience de l'utilisateur et une prise en compte de son avis.

Avec cette modélisation, nous mettons ainsi l'accent sur les actions de capitalisation que l'utilisateur peut avoir sur les deux entités présentées. Nous allons à présent explorer et implémenter les interactions permettant de réaliser ces actions.

4 Implémentation des interactions dans un prototype informatique

4.1 Identification des interactions pour la capitalisation

Capitaliser les indicateurs d'apprentissage résulte de leurs interactions avec les utilisateurs. Nous avons mené une étude sur des processus de conception des indicateurs [8] permettant d'examiner le rapport des utilisateurs non experts avec les indicateurs. Notre proposition se positionne dans ce lien entre les utilisateurs et indicateurs. Elle a donc nécessité d'être testée auprès d'utilisateurs pour observer, dans une démarche exploratoire au cours de plusieurs activités proposées, leurs interactions avec les indicateurs [9]. Les 47 participants se sont portés volontaires au sein du projet écri+ pour former 13 *focus groups* d'enseignants et étudiants. La notion d'indicateurs ne leur était pas forcément familière, mais ils ont pu les manipuler grâce aux tâches successives proposées et à l'interaction du groupe. Les actions ne se limitent pas à la conception de l'indicateur, mais se poursuivent dans l'utilisation qui en est faite. Il faut noter que ce n'est pas l'impact de l'indicateur ni son utilisation dans le cadre de l'apprentissage qui nous intéresse ici. L'efficacité de l'indicateur conçu par les utilisateurs n'est pas notre objet d'étude. Cette donnée est récoltée et proposée à l'utilisateur dans le cadre de l'accompagnement à la capitalisation, par exemple, comme critère de sélection d'un indicateur pour son usage.
C'est l'action combinée de plusieurs individus, voire de toute une communauté qui permet de capitaliser les indicateurs. La responsabilité ne reposant pas sur un seul individu, l'enjeu n'est plus d'avoir des utilisateurs fortement engagés dans le processus, mais plutôt que chacun puisse réaliser les actions qui lui sont pertinentes et avec lesquelles il se sent en confiance. Le partage d'expérience et ressenti permet une meilleure capitalisation en facilitant les interactions avec des indicateurs rendus plus concrets par la pratique et ainsi la création d'une communauté autour des indicateurs et leurs usages [21].
L'identification des interactions nécessaires aux utilisateurs pour capitaliser permet de proposer un système conçu pour la capitalisation.

4.2 Implémentation dans un prototype informatique

À partir du modèle d'indicateurs capitalisés présenté en partie 3, nous avons implémenté un prototype sous la forme d'une plateforme web. POPCORN (*Pro-*

totype fOr People's Capitalisation Of leaRning iNdicators), permettant aux utilisateurs d'expérimenter les interactions que nous avons identifiées pour la réalisation de la capitalisation d'indicateurs. Ce prototype vise dans un premier temps à proposer à des utilisateurs non experts, d'interagir avec des indicateurs d'apprentissage selon des modalités flexibles. Ci-après, nous présentons les différentes interactions implémentées regroupées dans les quatre actions de la capitalisation, à savoir la réutilisation, l'appropriation, l'adaptation et le partage.

Réutilisation Pour permettre la capitalisation, il faut commencer par rendre disponibles les indicateurs capitalisés dans le système. Pour cela, la fonctionnalité principale, et celle vers laquelle sont guidés les nouveaux utilisateurs, est d'explorer une bibliothèque d'indicateurs (voir Fig. 3). Ici les utilisateurs peuvent voir une description de tous les indicateurs existants, les informations qui leur sont liées et un aperçu des usages antérieurs. Il est possible de filtrer les indicateurs selon les différentes données typées qui leur sont attribuées, puis par recherche textuelle pour les champs libres, tels que le nom, la description et les commentaires.

Fig. 3. Capture d'écran de la page Bibliothèque de POPCORN, proposant des indicateurs comme un point de départ à la capitalisation

Sur cette page, et avec les informations disponibles, l'utilisateur peut choisir un *Indicateur*, et donc un *Usage* correspondant sera créé afin de permettre les interactions. L'utilisateur est ensuite dirigé vers un parcours d'appropriation.

Appropriation Une fois un indicateur choisi, l'utilisateur est invité à compléter des informations sur son objectif d'observation et son contexte pédagogique tels

que détaillés en partie 3.1, afin d'affiner sa compréhension de l'indicateur, et ainsi commencer l'appropriation de ce dernier (voir Fig. 4). Ces étapes amè-

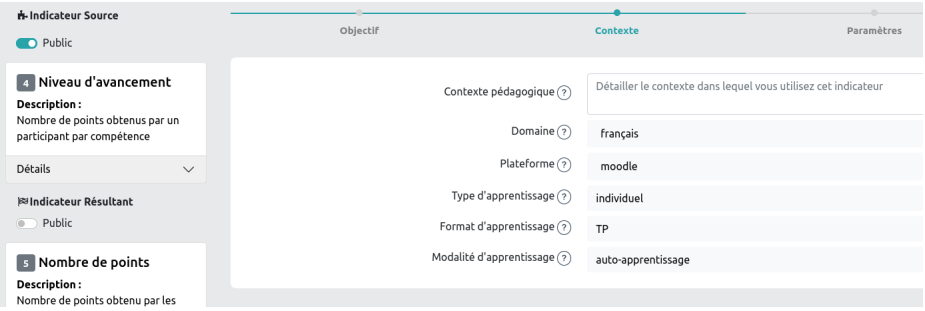


Fig. 4. Capture d'écran du parcours d'appropriation tel que scénarisé dans POPCORN

nent à un paramétrage nécessaire de l'indicateur pour son calcul pour l'usage de l'utilisateur. Ainsi, le processus d'analyse de l'indicateur est utilisé avec les données et les paramètres de l'utilisateur et le résultat conservé dans le cadre de son *Usage*. Il est également possible de modifier la visualisation de l'indicateur, en suivant ou non les recommandations du concepteur, sans que l'*Indicateur* en soit altéré.

Adaptation Les paramètres proposés ne sont pas toujours suffisants pour le besoin et contexte de l'utilisateur. Une étape d'adaptation peut alors être nécessaire. Plus technique, cette dernière est laissée au choix de l'utilisateur, le faire lui-même ou bien faire appel à un expert. Dans ce second scénario, notre objectif est de faciliter la communication entre l'utilisateur qui a un besoin et l'expert qui a des connaissances. L'interface alors proposée permet cette spécification de façon peu contraignante, mais complète, afin de minimiser les itérations. Lors de l'adaptation, toute proposition de modification du processus d'analyse de l'indicateur se fait à partir des informations déjà spécifiées, et non "de zéro". Cette modification, en impactant la partie *Processus* de l'*Indicateur*, et non plus son *Usage*, résulte en la création d'un nouvel indicateur, qui sera lié à l'usage de l'utilisateur ayant eu ce nouveau besoin.

Partage Un utilisateur peut également interagir avec d'autres, individus ou groupe, au travers des indicateurs qu'il a choisis. Le partage de l'indicateur correspond dans notre modèle à un partage de l'*Usage* d'un indicateur, et donc du contexte et de l'objectif personnalisé par l'utilisateur. Ainsi, un enseignant peut partager à ses collèges, comme à ses étudiants, son usage d'un indicateur. Le "receveur" pourra à son tour choisir l'indicateur tout en conservant les spécifications de l'usage de l'"envoyeur". Une autre interaction possible, bien que

moins directe, est l'évaluation. Elle a un double objectif, d'une part permettre un retour d'expérience aux prochains utilisateurs lors de leur choix, d'autre part avoir un critère de sélection plus automatique pour le système. Dans une version ultérieure, POPCORN pourra ainsi détecter, les indicateurs à "faible capitalisabilité", c'est-à-dire ne plaisant pas ou n'étant pas pertinents.

5 Expérimentation du prototype par les utilisateurs

Nous avons analysé le prototype réalisé pour vérifier qu'il remplit les objectifs fixés lors de sa conception. Comment valider notre approche de capitalisation par les interactions ? Le prototype réalisé nous donne un support permettant d'établir un protocole pour expérimenter la capitalisation par les utilisateurs.

5.1 Protocole de test utilisateur

Dans le cadre de la mise en place de test utilisateur, nous avons réduit le contexte du prototype par rapport au projet écri+ auquel il est lié. Ainsi, les indicateurs proposés par le prototype viennent directement des *focus group* mentionné en section 4.1. De même, les 7 participants au test du prototype sont en priorité des membres, aux profils majoritairement enseignants, de ce même projet. Ces enseignants sont, bien que peu habitués à la manipulation d'indicateurs, relativement expérimentés dans les pratiques pédagogiques. Le format choisi pour le test utilisateur est celui d'entretien individuel scénarisé incluant une manipulation du prototype POPCORN présenté dans la section précédente. L'expérimentation permet la collecte d'informations sous plusieurs formats, notes, enregistrements et traces d'activités, dans le cadre d'une démarche exploratoire. De plus, l'expérimentation permet de créer des *Usages* propres à chaque participant pour les *Indicateurs* choisis.

5.2 Objectif de l'expérimentation

L'objectif d'un test impliquant des utilisateurs non experts est d'analyser leur utilisation et leur intérêt pour le système d'accompagnement vers la capitalisation. Le but est d'observer si les actions réalisées par les utilisateurs vont dans le sens d'une capitalisation de qualité, et si les interactions proposées leur semblent naturelles. Les critères caractérisant cette *qualité* de la capitalisation sont un point que nous travaillons actuellement. Nous tiendrons compte de l'analyse de la production de l'utilisateur, mais aussi des interactions, collectées à l'aide des sondes d'activités placées dans le prototype, et enfin de l'évolution de l'indicateur et de ces usages dans le temps. Ce dernier élément n'est actuellement pas récolté dans le cadre d'une expérimentation limitée dans le temps.

5.3 Premières itérations du test utilisateurs & perspectives

À la suite des premières itérations du test utilisateurs, plusieurs observations peuvent déjà être soulignées, bien que nécessitant une validation ultérieure :

- La complexité perçue des interactions est faible, telle qu'observée au travers des verbatims récoltés. Il faut en revanche noter que le prototype n'implémente pas encore un certain nombre d'interactions plus complexes (comme une partie du paramétrage).
- La partie visualisation des indicateurs n'a pas été détaillée dans cet article, très appréciée des utilisateurs, c'est une fonctionnalité nécessaire dans l'accompagnement à la capitalisation.
- À l'instar des choix faits dans le cadre du test, il nous semble pertinent de garder une instance spécialisée pour un EIAH, où une communauté de pratique pourra s'emparer des indicateurs et des fonctionnalités proposés par POPCORN.
- Une nouvelle version de la proposition de modélisation peut être complétée avec des mécanismes supplémentaires d'interactions entre les entités, comme la combinaison de plusieurs indicateurs.
- Pour guider les utilisateurs de façon plus efficace, l'automatisation d'une partie des tâches, est une fonctionnalité qui nécessite une base structurée d'indicateurs capitalisés et des informations sur les interactions des utilisateurs pour la capitalisation. Les expérimentations ont montré que les participants sont demandeurs de ce type d'accompagnement.

6 Conclusion

L'utilisation d'indicateurs d'apprentissage dans le cadre des EIAH est à destination des acteurs de l'apprentissage dont la diversité et l'évolution constante des pratiques sont peu compatibles avec la complexité de conception d'indicateurs pertinents. La proposition d'une approche de capitalisation amène à une valorisation de l'implication des utilisateurs. Leurs interactions avec les indicateurs permettent de faire évoluer ces derniers au cours des usages. Cette approche nécessite tout d'abord une modélisation des indicateurs capitalisés, incluant la multiplicité des usages par les utilisateurs. Une fois les entités créées, nous avons implémenté les interactions permettant aux utilisateurs de capitaliser les indicateurs. La réalisation d'un premier prototype sous forme de plateforme web permet d'observer les utilisateurs dans des situations de capitalisation des indicateurs et de se diriger vers un meilleur accompagnement des différentes interactions identifiées. Cette démarche exploratoire de l'accompagnement à la capitalisation est prometteuse dans son accueil par les utilisateurs. Les tests initialement mis en place vont permettre de planifier les prochaines évolutions, notamment pour mettre les utilisateurs non experts en confiance tout en répondant aux besoins techniques de la capitalisation.

Acknowledgements Notre travail est mené dans le cadre du projet PIA écri+ (n° ANR-17-NCUN- 0015) <http://ecriplus.fr/>.

References

1. Alowayr, A., Badii, A.: Review of monitoring tools for e-learning platforms. IJCSIT (2014)

2. Basak, S.K., Wotto, M., Bélanger, P.: A framework on the critical success factors of e-learning implementation in higher education: A review of the literature. International journal of educational and pedagogical sciences **10**(7), 2409–2414 (2016)
3. Cherigny, F., El Kechai, H., Iksal, S., Lefevre, M., Labarthe, H., Luengo, V.: L'analytique des apprentissages avec le numérique groupes thématiques de la direction du numérique pour l'éducation (dne-tn2). Direction du numérique pour l'éducation (2020)
4. Dabbebi, I., Gilliot, J.M., Iksal, S.: User centered approach for learning analytics dashboard generation. In: CSEDU. vol. 2, pp. 260–267 (2019)
5. Diagne, F.: Instrumentation de la supervision par la réutilisation d'indicateurs: Modèles et Architecture. Ph.D. thesis, Université Joseph-Fourier-Grenoble I (2009)
6. Dimitrakopoulou, A.: State of the art on interaction and collaboration analysis (2004)
7. Dollinger, M., Liu, D., Arthars, N., Lodge, J.M.: Working together in learning analytics towards the co-creation of value. Journal of Learning Analytics **6**(2), 10–26 (2019)
8. Gril, A., May, M., Renault, V., George, S.: A comparative analysis of approaches to design and capitalize data indicators. In: Open and Inclusive Educational Practice in the Digital World, pp. 135–151. Springer (2022)
9. Gril, A., Renault, V., May, M., George, S.: Learning indicator capitalization: Protocol for needs gathering and user-centered design. pp. 71–73. IEEE (2022)
10. Hoffmann, C., Mandran, N., d'Ham, C., Rebaudo, S., Haddouche, M.A.: Development of actionable insights for regulating students' collaborative writing of scientific texts. In: EC-TEL. pp. 534–541. Springer (2022)
11. Iksal, S.: Ingénierie de l'observation basée sur la prescription en EIAH. Ph.D. thesis, Université du Maine (2012)
12. Inventado, P.S., Scupelli, P.: An online learning collaboratory to address multidisciplinary learning challenges at scale. IxD&A **33**, 11–32 (2017)
13. Jørnø, R.L., Gynther, K.: What constitutes an 'actionable insight'in learning analytics? Journal of Learning Analytics **5**(3), 198–221 (2018)
14. Lebis, A., Lefevre, M., Luengo, V., Guin, N.: Capitalisation of analysis processes: Enabling reproducibility, openness and adaptability thanks to narration. In: LAK. pp. 245–254 (2018)
15. Martinez-Maldonado, R., Pardo, A., Mirriahi, N., Yacef, K., Kay, J., Clayphan, A.: Latux: An iterative workflow for designing, validating, and deploying learning analytics visualizations. Journal of Learning Analytics **2**(3), 9–39 (2015)
16. May, M., Iksal, S., Usener, C.A.: The side effect of learning analytics: An empirical study on e-learning technologies and user privacy. In: International Conference on Computer Supported Education. pp. 279–295. Springer (2016)
17. Michel, C., Lavoué, E., George, S., Ji, M.: Supporting awareness and self-regulation in project-based learning through personalised dashboards. IJTEL (2017)
18. Ollagnier-Beldame, M.: Foot-printing e-learners' activity: A first step to help their appropriation of the training system? IJKM **6**(3), 65–77 (2010)
19. Rodríguez-Triana, M.J., Prieto, L.P., Vozniuk, A., Boroujeni, M.S., Schwendimann, B.A., Holzer, A., Gillet, D.: Monitoring, awareness and reflection in blended technology enhanced learning: a systematic review. IJTEL **9**(2-3), 126–150 (2017)
20. Sedrakyan, G., Mannens, E., Verbert, K.: Guiding the choice of learning dashboard visualizations: Linking dashboard design and data visualization concepts. Journal of Computer Languages **50**, 19–38 (2019)
21. Soliman, D., Costa, S., Scardamalia, M.: Knowledge building in online mode: Insights and reflections. Education Sciences **11**(8), 425 (2021)

Caractérisation et mesure des discriminations algorithmiques dans la prédiction de la réussite à des cours en ligne

Mélina Verger, François Bouchet, Sébastien Lallé et Vanda Luengo

Sorbonne Université, CNRS, LIP6, F-75005 Paris, France

Résumé. Les modèles prédictifs utilisés en EIAH peuvent produire des résultats biaisés et discriminants. Or, les mesures existantes pour les détecter sont seulement fondées sur l'égalité des performances prédictives entre différents groupes d'apprenant·e·s. Dans cet article, nous proposons une mesure objective des discriminations algorithmiques d'un modèle, ainsi qu'une méthode d'analyse visuelle pour caractériser ces discriminations. Nous démontrons l'application de notre méthode dans le cadre de la prédiction de la réussite à des cours en ligne, au moyen de données éducatives ouvertes. Nos résultats montrent la nécessité d'analyser systématiquement les discriminations algorithmiques issues des modèles afin de confirmer ou d'infirmer le caractère sensible de certains attributs.

Mots-clés : Équité algorithmique , métrique , attributs sensibles

Abstract. Preditive models used in intelligent learning environments can suffer from biased and unfair representation. However, existing fairness metrics that are meant to capture these issues are only based on the models' predictive performances. In this paper, we propose a novel fairness metric that measures to what extent the models behave unfairly. In addition, we provide a visualization-based analysis to qualify the types of unfair behaviors that are exhibited by the models. We apply our method on the success prediction task in online courses, with an open educational dataset. Our results highlight the need to systematically analyze unfair behaviors from the models in order to confirm or refute the sensitive nature of some attributes.

Keywords: Algorithmic fairness , metric , sensitive attributes

1 Introduction

La prédiction de la réussite à des cours en ligne a connu un intérêt important ces dix dernières années, comme le montre notamment une revue systématique relevant 357 articles publiés entre 2010 et mi-2018 sur le sujet [4]. Cependant, plusieurs exemples ont montré que les modèles prédictifs, au-delà des inégalités

et discriminations déjà présentes dans la réalité, pouvaient aussi être biaisés en faveur ou au détriment de certains groupes de la population, et ainsi amplifier voire produire de nouvelles discriminations [9,2,3,8]. Pour les distinguer des *discriminations historiques*, qui sont celles déjà présentes dans la société, nous nous réfèrerons à ces nouvelles discriminations issues de l'utilisation des modèles prédictifs sous le terme de *discriminations algorithmiques*. Ce constat alarmant sur l'utilisation des modèles prédictifs a conduit à une prise de conscience de leurs risques potentiels et notamment à la mise en place de directives par les instances de régulation [1].

Ainsi, il est devenu nécessaire d'analyser les discriminations apprises et produites par les modèles prédictifs. En particulier, une telle analyse doit permettre d'identifier envers quels groupes spécifiquement les modèles produisent les résultats les plus biaisés, pour pouvoir appréhender les implications réelles de l'utilisation de ces modèles. Dans la littérature (voir section 2), l'approche systématiquement employée pour quantifier les discriminations algorithmiques d'un modèle consiste à identifier au préalable les groupes de personnes potentiellement concernés, puis de comparer les performances prédictives (e.g. précision, F1-score) du modèle en fonction de ces groupes distincts. Un modèle est alors considéré comme ayant un comportement discriminant envers un groupe s'il ne produit pas des performances prédicitives similaires à celles obtenues pour les autres groupes.

Pour autant, évaluer si un modèle produit des performances prédictives similaires entre différents groupes (i.e. le même nombre d'erreurs) ne tient pas compte du fait qu'il puisse produire des erreurs plus sévères (et dont les implications seraient plus néfastes) pour un groupe que pour un autre, en dépit d'un nombre d'erreurs identique. C'est pourquoi nous proposons une nouvelle méthode d'analyse des discriminations algorithmiques visant à quantifier la sévérité de ces erreurs grâce à une nouvelle métrique nommée *Model Absolute Density Distance* (MADD) et fondée sur la caractérisation de ces discriminations via une analyse visuelle dérivée de cette métrique. Cette nouvelle métrique et l'analyse visuelle associée sont indépendantes des performances prédictives des modèles, pour permettre de quantifier les discriminations algorithmiques uniquement. Cette approche est particulièrement destinée aux chercheurs·euses et développeurs·euses de modèles prédictifs en éducation.

Nous appliquons notre approche sur un cas d'usage de prédiction de la réussite à des cours en ligne avec des données éducatives ouvertes, dans un souci de reproductibilité des expériences, et avec quatre types de modèles de classification binaire très courants pour cette tâche [1], dans un souci de généralisation. Nous mettons à disposition les données et le code documenté, permettant la réplication et l'utilisation de notre méthode dans d'autres contextes, à l'adresse suivante : `https://github.com/melinaverger/MADD`.

1. Règlement Général sur la Protection des Données (2016) au niveau européen, *California Consumer Privacy Act* (2018) au niveau des États-Unis, Principes de l'OCDE (Organisation de coopération et de développement économiques) sur l'intelligence artificielle (2019) au niveau international, et prochainement l'*Artificial Intelligence Act*.

2 État de l'art

D'après différentes revues de littérature [4,6], la prédiction de la réussite à des cours en ligne est le plus souvent représentée par un problème de classification binaire (e.g. réussite/échec). Plusieurs types de modèles sont couramment utilisés, tels que les réseaux bayésiens, les arbres de décisions, ou la méthode des k plus proches voisins. En expérimentant notre approche avec quatre types de modèles de classification binaire très courants en éducation [4,6] et en particulier sur le corpus OULAD [7,1], notre contribution s'inscrit dans la continuité de ces pratiques, et nous détaillerons ces choix dans la section 4. Le corpus OULAD a notamment été utilisé dans plusieurs travaux de prédiction connexes (i.e. réussite/échec, abandon/complétion des cours) [1], mais sans analyse des discriminations algorithmiques dans les prédictions.

Par ailleurs, quelques travaux en éducation [3,5,8], menés aux Etats-Unis, ont cherché à évaluer les discriminations apprises par des modèles de classification binaire. Or, comme abordé en introduction, au lieu de déterminer les groupes d'apprenant·e·s subissant des discriminations algorithmiques à partir des prédictions des modèles, les auteurs choisissent d'abord les groupes d'apprenant·e·s qu'ils jugent à risque, puis comparent les performances des modèles entre ces différents groupes. [3] les a par exemple comparé par rapport au genre, et [5] et [8] par rapport au genre et à l'origine ethnique des apprenant·e·s [2]. Le choix préalable de ces caractéristiques à risque, aussi appelées *attributs sensibles*, est souvent basé sur les résultats d'études de sciences sociales ou sur les caractéristiques mises en avant par des lois anti-discrimination (e.g. genre, origine ethnique, religion, handicap). En revanche, cette approche ne peut pas rendre visible les potentielles discriminations algorithmiques par rapport à d'autres attributs sensibles non sélectionnés par les auteurs. En effet, les discriminations apprises par un modèle peuvent être différentes de celles présupposées, puisqu'elles dépendent non seulement de la nature et de la représentation des données utilisées, mais aussi de la manière dont le modèle apprend de celles-ci. C'est pourquoi nous proposons une nouvelle approche fondée à la fois sur la quantification et la caractérisation des discriminations algorithmiques (voir section 3), afin de confirmer ou d'infirmer *a posteriori* le caractère *sensible* de certains attributs.

Enfin, notre approche repose exclusivement sur l'analyse des discriminations algorithmiques des modèles indépendamment de leurs performances prédictives, et diffère en cela des travaux existants sur l'équité des modèles prédictifs en éducation [3,5,8]. En effet, ces travaux comparent par exemple la différence de précision des modèles entre les différents groupes, ou le taux de bonnes prédictions par rapport au taux de mauvaises prédictions. Cependant, évaluer si ces performances prédictives sont égales à travers les groupes n'est pas synonyme d'absence de discrimination : un modèle peut produire des erreurs en même quantité, mais qui peuvent être très nuisibles pour un groupe et très peu pour

2. Dans l'Union Européenne (UE), les analyses sur l'origine ethnique ne peuvent être conduites du fait du Règlement Général sur la Protection des Données (RGPD) interdisant la collecte de ce type d'information.

l'autre. Dans la mesure où différentes formes de discrimination algorithmique existent [12,10], nous avons développé une mesure indépendante de la performance prédictive, la MADD, capable de les quantifier.

3 Méthode d'analyse des discriminations algorithmiques

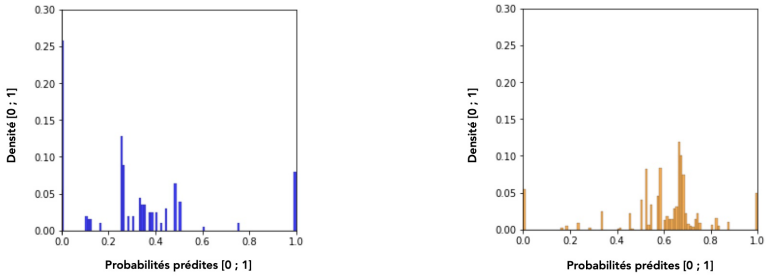
Dans cette section, nous présentons notre méthode d'analyse des discriminations algorithmiques des modèles prédictifs de classification binaire, tout d'abord via l'explication de l'analyse visuelle qui la compose (partie 3.1) puis via la définition de la nouvelle métrique MADD (partie 3.2).

Avant cela, considérons des modèles de classification binaire pour la prédiction la réussite à un cours en ligne. Pour appliquer notre méthode, chaque modèle doit fournir pour chaque prédiction (i.e. 0 pour échec ou 1 pour réussite) soit une estimation de sa probabilité pour les modèles probabilistes (e.g. réseaux bayésiens) soit un score de confiance pour les modèles non probabilistes (e.g. arbres de décision), les deux étant représentés par une valeur comprise entre 0 et 1. Par simplification, nous utiliserons les termes *probabilités prédites* ou *probabilités* pour faire référence à la fois aux estimations de probabilité et aux scores de confiance. Par exemple, avec un seuil de classification fixé à 0,5, un modèle prédit la valeur 1 (réussite) s'il produit une *probabilité* supérieure à 0,5, et prédit 0 (échec) sinon.

3.1 Analyse visuelle des discriminations algorithmiques

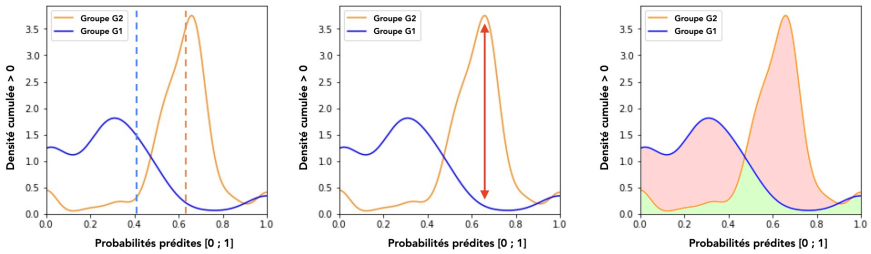
Au lieu de nous intéresser seulement aux prédictions 0 ou 1 que produisent les modèles, comme comparées dans les travaux cités en section 2, nous étudions de manière plus fine leurs probabilités prédites. Pour cela, nous étudions les fréquences avec lesquelles les modèles attribuent ces probabilités, en particulier celles associées à la prédiction 1 (réussite). Par exemple, dans la Figure 1, les histogrammes montrent pour un modèle donné la distribution des probabilités liées à la réussite pour deux groupes d'apprenant·e·s distincts, G1 et G2 (e.g. les apprenant·e·s déclarés avec un handicap (G1) et les apprenant·e·s déclarés sans handicap (G2)). Chaque barre verticale sur les Figures 1a et 1b représente la proportion d'apprenant·e·s ayant reçue la même probabilité de réussite. Nous appellerons par la suite une telle distribution *vecteur de densité* des probabilités.

A titre d'exemple, sur ces histogrammes nous pouvons constater que les probabilités de G1 sont surtout situées entre 0 et 0,5, alors que celles de G2 sont plus élevées, principalement entre 0.5 et 0.7 environ. Le modèle a donc tendance à donner de meilleures probabilités de réussite à G2 qu'à G1. Ainsi, pour faciliter l'analyse visuelle de ces histogrammes, difficilement interprétables en raison des nombreuses variations observées, nous proposons d'appliquer un lissage par une méthode d'estimation de densité par noyau (ou *kernel density estimation*). Nous utilisons plusieurs noyaux gaussiens pour approximer les distributions discrètes montrées dans les histogrammes et le coefficient de lissage est calculé automati-



(a) Groupe G1 (b) Groupe G2

FIG. 1 – Histogrammes des probabilités prédites pour deux groupes distincts.



(a) Traitement inégal (b) Jugement stéréotypé (c) MADD approximée

FIG. 2 – Représentations visuelles des formes de discriminations algorithmiques (a, b) et de la MADD (c). Les lignes en pointillées représentent les moyennes des distributions de probabilités.

quement par la règle de Scott [3] qui prend en compte le nombre d'échantillons et le nombre d'attributs présents dans le jeu de données. La Figure 2 présente le résultat d'un tel lissage pour les histogrammes de la Figure 1. Le passage de la distribution discrète à la distribution lissée change la densité en densité cumulée (axe des ordonnées), ce qui explique pourquoi les valeurs observées peuvent être supérieures à 1 dans la Figure 2.

Ce lissage permet de caractériser deux formes de discriminations algorithmiques :

1. Le *traitement inégal* : un modèle peut donner en moyenne de meilleures probabilités à un groupe qu'à un autre (Figure 2a), ce qui traduit le favoritisme du modèle pour un groupe par rapport à l'autre.
2. Le *jugement stéréotypé* : un modèle peut donner à de nombreux apprenant·e·s dans un groupe la même probabilité, ce qui traduit un comportement répétitif et invariant, caractéristique du stéréotype (Figure 2b).

Nous cherchons donc, dans la partie 3.2 suivante, à quantifier ce qui correspond à la zone rouge en Figure 2c, zone où un modèle ne produit pas les mêmes

3. Voir la documentation Python de `scipy.stats.gaussian_kde`.

probabilités de réussite pour les deux groupes. Il est cependant important de noter que cette quantification n'utilise pas les résultats du lissage effectué, qui ne vise qu'à faciliter l'analyse visuelle, mais bien les probabilités qui sont effectivement prédites par les modèles.

3.2 Métrique *Model Absolute Density Distance* (MADD)

Nous définissons la métrique MADD [4] comme suit. Posons les vecteurs de densité $D^{G1} = [d_0^{G1}, d_1^{G1}, ..., d_m^{G1}]$ et $D^{G2} = [d_0^{G2}, d_1^{G2}, ..., d_m^{G2}]$ associés aux groupes G1 et G2 respectivement, où m correspond au nombre de valeurs discrètes possibles que les probabilités de réussite peuvent prendre [11]. Comme chaque vecteur représente la fréquence des probabilités, la somme de ses éléments vaut toujours 1. Ainsi :

$$\text{MADD}(D^{G1}, D^{G2}) = \sum_{k=0}^{m} |d_k^{G1} - d_k^{G2}| \tag{1}$$

La MADD est bornée entre 0 et 2. En effet, la MADD vaut 0 quand les deux vecteurs de densité sont identiques, c'est-à-dire que le modèle a le même comportement pour G1 et G2. A l'inverse, la MADD vaut 2 quand le modèle ne produit aucune probabilité commune entre les deux groupes. Une telle situation se produit par exemple quand le modèle donne une probabilité unique de p_i à tous les apprenant·e·s de G1 (i.e. densité maximale pour une seule valeur de probabilité donnée) et une probabilité de p_j (avec $p_j \neq p_i$) à tous les apprenant·e·s de G2. Ainsi, pour n'importe quelles probabilités données, indexées par i et j :

$$\text{MADD}(D^{G1}, D^{G2}) = |d_i^{G1}| + |d_j^{G2}| = (1 + 1) = 2 \tag{2}$$

4 Expériences

4.1 Corpus de données OULAD

Nous expérimentons notre méthode sur le jeu de données OULAD (*Open University Learning Analytics Dataset*) [7]. Il s'agit en effet d'un corpus anonymisé largement utilisé en éducation [1], y compris pour la prédiction de la réussite à des cours en ligne ; les données sont ouvertes, répondant spécifiquement à l'appel lancé à la communauté pour le développement de nouvelles approches sur des jeux de données ouverts [4] ; et il contient des données de différents cours avec des profils d'apprenant·e·s variés, ce qui nous permet de répliquer nos expériences dans plusieurs contextes avec des populations différentes (autre appel lancé par [4]). De plus, les données ont été collectées avec une attention particulière sur l'éthique et le respect de la vie privée.

Les cours du jeu de données OULAD ont été dispensés par *The Open University*, une université britannique à distance qui propose des cours pouvant être

4. Traduisible en "Distance Absolue entre les Densités du Modèle".

Table 1 – Attributs utilisés du jeu de données OULAD.

Attribut	Type	Description
`genre`	binaire	genre de l'apprenant·e
`age`	ordinal	intervalle de l'âge de l'apprenant·e
`handicap`	binaire	indique si l'apprenant·e a déclaré un handicap
`dernier_diplome`	ordinal	dernier diplôme de l'apprenant·e en entrée du cours
`pauvrete`	ordinal	niveau de pauvreté du lieu d'habitation de l'apprenant·e
`nb_tentatives`	numérique	nombre de tentatives précédentes au cours
`credits`	numérique	nombre de crédits pour le cours étudié par l'apprenant·e
`nb_total_click`	numérique	nombre total d'interactions de l'apprenant·e avec le cours

suivis sans prérequis de manière indépendante ou dans le cadre d'un cursus universitaire. Les apprenant·e·s étaient inscrit·e·s entre 2013 et 2014 à au moins un des sept cours recensés dans le OULAD, dont trois en sciences sociales et quatre en Science, Technologie, Ingénierie et Mathématiques (STIM).

Le corpus contient des données démographiques et des données d'activité dans l'espace numérique de travail (ENT), avec 28 785 échantillons (paire apprenant·e - cours). Nous avons utilisé les attributs présentés en Table 1 ainsi que la variable cible binaire "Réussite"/"Echec". Seul l'attribut `nb_total_click` n'était pas immédiatement disponible dans le corpus et a été calculé par jointure et agrégation. Nous avons supprimé les échantillons avec des données manquantes et les valeurs de chaque attribut ont été normalisées entre 0 et 1 en prenant soin de ne pas appliquer de standardisation précisément pour garder les distributions de données originales pour l'analyse des discriminations algorithmiques.

4.2 Attributs sensibles d'étude et sélection des cours

Pour nos expériences, nous ciblons l'étude du caractère sensible aux quatre attributs suivants : `genre`, `age`, `pauvrete` et `handicap`. Dans une recherche exhaustive d'attributs sensibles, il est tout à fait possible d'analyser avec notre méthode les discriminations algorithmiques relativement à tous les attributs disponibles dans un jeu de données. Par exemple, les attributs `dernier_diplome`, `nb_tentatives`, `credits` et `nb_total_click` pourraient être pris en compte de manière complémentaire pour évaluer, en plus d'informer respectivement sur l'état des connaissances préalables, l'expérience du cours, son attractivité et l'intensité de l'activité des apprenant·e·s en son sein qui ne sont pas des informations à risque pour la prédiction du succès, leur sensibilité sans être démographiques. Nous allons ici plutôt confirmer ou infirmer le caractère sensible des quatre attributs retenus. Par ailleurs, pour distinguer deux groupes G1 et G2 pour `pauvrete` et `age` respectivement, nous utilisons un seuil de 50% de l'indice de pauvreté britannique (voir [7]) et nous distinguons, parmi les trois tranches d'âge disponibles dans les données ([0-35], [35-55] et [55+]), le groupe majoritaire ([0-35]) du groupe minoritaire (regroupement de [35-55] et [55+]).

Quant aux cours étudiés, nous avons sélectionné un cours en sciences sociales, identifié "BBB" dans le corpus, et un cours en STIM, identifié "FFF". En effet,

d'après une analyse des corrélations des attributs, ces deux cours ont présenté les plus fortes corrélations avec l'attribut `genre`, ce qui suggère une importance de cet attribut pour la prédiction de la réussite ou de l'échec par les modèles. De plus, ces deux cours ont aussi présenté de forts déséquilibres entre les deux groupes le constituant dans le corpus, c'est-à-dire une large majorité de femmes (91.2 %) dans le cours "BBB", et à l'inverse une majorité d'hommes dans le cours "FFF" (88.4 %). Le choix de ces deux cours était donc pertinent pour notre analyse des discriminations algorithmiques par rapport aux attendus de biais de genre.

4.3 Modèles prédictifs de la réussite

Dans un souci de généralisation, nous expérimentons notre approche avec plusieurs types de modèles de classification, respectivement à base de régression, de distances, d'arbres et de probabilités : un modèle de régression logistique (LR), un modèle des k-plus proches voisins (KN), un arbre de décision (DT) et un classifieur naïf bayésien (NB). Le choix de ces modèles a été motivé par plusieurs raisons. Tout d'abord, les modèles susmentionnés sont largement utilisés dans le domaine de l'éducation [6,1] et y compris avec le jeu de données OULAD (voir section 2). D'autres modèles courants comme les machines à vecteurs de support n'ont pas été retenus car ils ne produisent pas d'estimations de probabilité (ou de scores de confiance) nécessaires pour effectuer notre analyse. Deuxièmement, bien que notre approche puisse être généralisée à d'autres modèles tels que des forêts aléatoires et des réseaux de neurones, nous avons privilégié les boîtes blanches et l'explicabilité sur l'optimisation que requiert ces modèles. Troisièmement, la prédiction de la réussite avec les données OULAD est un problème de prédiction à faible niveau d'abstraction, où l'utilisation de modèles prédictifs complexes conduirait à de moins bonnes performances et à un surapprentissage.

Nous avons entraîné les modèles en utilisant 70% des données pour le jeu d'entraînement et 30% pour le jeu de test, en gardant les mêmes proportions de réussite et d'échec dans les deux jeux. Les modèles ont obtenu des précisions supérieures à la précision de référence (70% étant la proportion originale de réussite) allant jusqu'à 93%, à l'exception du classifieur NB (62%) qui en revanche a présenté des comportements intéressants pour l'analyse des discriminations algorithmiques. Nous soulignons à nouveau que contrairement aux études d'apprentissage automatique classiques, l'objectif ici n'est pas d'obtenir les meilleures performances prédictives mais de présenter l'intérêt de notre méthode sur divers modèles. Puis, nous avons calculé la MADD et réalisé les analyses visuelles sur le jeu de test.

5 Résultats

5.1 Cours de sciences sociales ("BBB")

Le Tableau 2 présente les résultats de la MADD pour chaque modèle et chaque attribut sensible dans le cours de sciences sociales. Les meilleurs résultats de

MADD par attribut (lecture en colonne) sont en gras, et les meilleurs résultats de MADD par modèle (lecture en ligne) portent une astérisque. Les valeurs les plus élevées représentant les discriminations les plus fortes par modèle sont en rouge. Le Tableau 2 montre ainsi que `handicap` est l'attribut vis-à-vis duquel trois modèles sur quatre (LR, KN et DT) discriminent le moins, avec la moyenne la plus basse à 0.82. À l'inverse, pour identifier les apprenant·e·s les plus discriminé·e·s, nous nous intéressons aux attributs qui induisent une valeur de MADD la plus élevée à travers tous les modèles. L'attribut `pauvrete` s'avère

TABLE 2 – Résultats de la MADD pour le cours "BBB".

	Modèle	Attributs sensibles				
		`genre`	`age`	`handicap`	`pauvrete`	Moyenne
MADD	LR	1.72	1.80	1.57*	1.86	1.74
	KN	1.13	1.12	0.93*	1.13	1.08
	DT	**0.69**	**0.84**	**0.65***	**0.85**	0.76
	NB	**0.69***	1.14	1.13	0.87	0.96
	Moyenne	1.06	1.23	0.82	1.18	

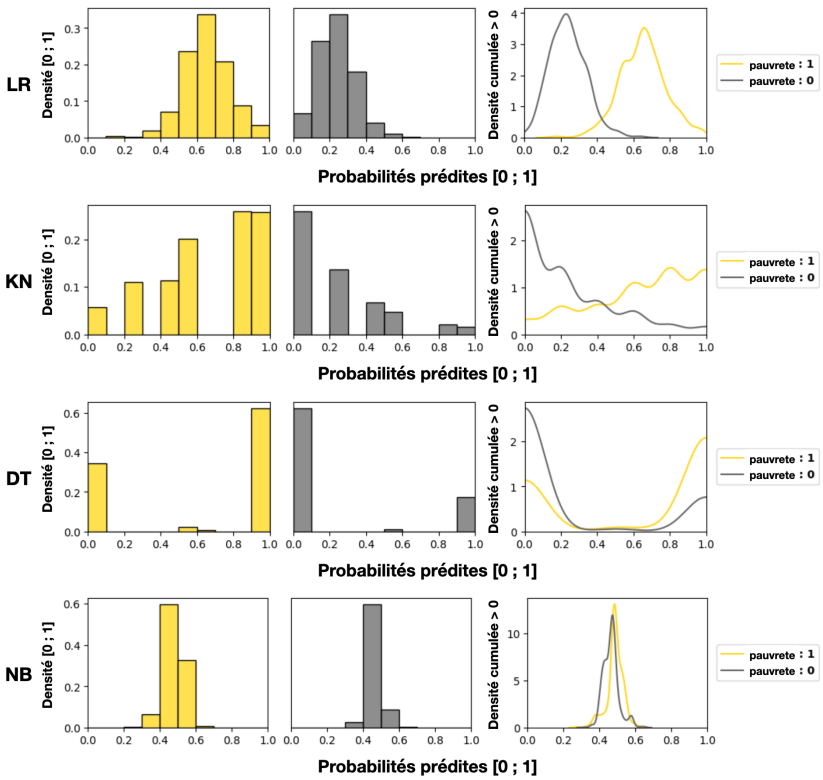


FIG. 3 – Analyse visuelle des vecteurs de densité pour le cours "BBB". Les barres des histogrammes ont été épaissies à des fins de visualisation.

être le plus discriminant pour trois modèles sur quatre (LR, KN et DT). Ainsi, pour caractériser les formes de discriminations algorithmiques vis-à-vis de cet attribut et identifier le groupe de `pauvretre` spécifiquement le plus discriminé, nous observons en Figure 3 que LR, KN et DT ont appris un "traitement inégal" (i.e. écart de moyennes significatif) en défaveur du groupe 0, les personnes vivant dans des régions les moins pauvres, obtenant les moins bonnes probabilités de réussite. Par conséquent, dans ce cours, les apprenant·e·s vivant dans des régions plus aisées sont les plus négativement discriminé·e·s (traitement inégal) par

TABLE 3 – Résultats de la MADD pour le cours "FFF".

	Modèle	Attributs sensibles				
		`genre`	`age`	`handicap`	`pauvrete`	Moyenne
MADD	LR	1.20	1.10	1.09	1.05*	1.11
	KN	1.05	0.96	0.79*	0.92	0.93
	DT	0.78	**0.68**	**0.60***	0.67	0.68
	NB	**0.53**	0.97	0.93	**0.44***	0.72
	Moyenne	0.89	0.93	0.85	0.77	

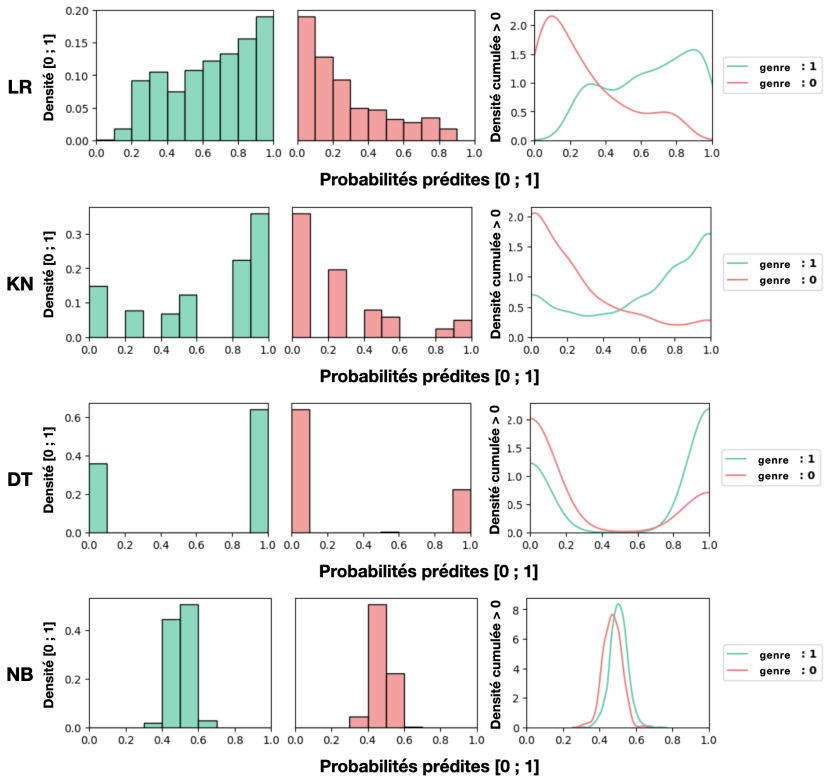


FIG. 4 – Analyse visuelle des vecteurs de densité pour le cours "FFF". Les barres des histogrammes ont été épaissies à des fins de visualisation.

la majorité des modèles, ou inversement ces modèles discriminent positivement les apprenant·e·s des régions défavorisées. Le NB a lui appris un comportement plus équilibré tandis que le DT et le KN montrent un comportement très stéréotypé avec seulement quelques pics distincts (voir histogrammes en Figure 3, qui montrent par exemple que pour le groupe 1, le KN donne 0.6, 0.8 ou 1.0 à 70 % des apprenant·e·s). Cela reflète bien le fonctionnement inhérent à ces deux modèles.

5.2 Cours de STIM (FFF)

Pour le cours de STIM, le Tableau 3 montre qu'en revanche `pauvrete`, avec la moyenne la plus basse (0.77), est l'attribut vis-à-vis duquel les modèles discriminent le moins. A l'inverse, l'attribut qui engendre le plus de discrimination pour trois modèles sur quatre est cette fois le `genre`, même si la moyenne pour l'`age` est également élevée. Ainsi, d'après la Figure 4, les comportements des modèles révèlent majoritairement une inégalité de traitement en défaveur du groupe 0, les femmes. Par conséquent, dans ce cours, les femmes représentent le groupe le plus négativement discriminé (traitement inégal) par les modèles.

6 Conclusion

Les résultats conduisent à deux conclusions principales. Premièrement, il n'y a pas de relation directe entre les biais dans les données étudiées en entrée et les biais dans les discriminations algorithmiques des modèles en sortie. Ainsi, malgré le biais de genre dans les données du cours de sciences sociales, notre analyse montre que c'est un autre attribut sensible, pauvreté, qui est a l'origine des discriminations algorithmiques les plus importantes. Deuxièmement, les analyses visuelles permettent de montrer que chaque modèle, même entraîné sur des données identiques, produit des discriminations algorithmiques différentes, observables à travers la variabilité des distributions dans les Figures 3 et 4.

Ainsi, ces conclusions démontrent la nécessité d'analyser systématiquement les discriminations algorithmiques des modèles prédictifs pour confirmer ou infirmer le caractère *sensible* de certains attributs. Cependant, il faut souligner qu'en pratique la MADD doit être utilisée avec des modèles présentant des performances prédictives satisfaisantes pour être utilisés dans des applications réelles. Dans notre cas, ceci exclurait le modèle NB qui, bien que présentant de bons résultats de MADD pour les attributs sensibles, en présentait aussi pour tous les attributs du corpus, le rendant par conséquent mauvais prédicteur de la réussite ou de l'échec (d'où sa plus faible précision en partie 4.3).

Notre méthode peut être utilisée de la même manière que dans cet article pour (1) quantifier les discriminations algorithmiques selon différents attributs, (2) caractériser la nature de ces discriminations, et (3) identifier spécifiquement les groupes les plus discriminés par les modèles, dans différents contextes. Notre travail vise plus largement à encourager la communauté à analyser les comportements des modèles actuellement intégrés dans des EIAH, et

nous mettons à disposition les données et le code documenté à l'adresse suivante : `https://github.com/melinaverger/MADD`. Nos prochains travaux visent à la généralisation de la définition de la MADD pour prendre en compte l'influence de plusieurs attributs simultanément au lieu d'un seul.

Références

1. Alhakbani, H.A., Alnassar, F.M. : Open Learning Analytics : A Systematic Review of Benchmark Studies using Open University Learning Analytics Dataset (OULAD). In : 7th International Conference on Machine Learning Technologies (ICMLT). pp. 81–86. ACM, New York, NY, USA (2022)
2. Buolamwini, J., Gebru, T. : Gender Shades : Intersectional Accuracy Disparities in Commercial Gender Classification. In : 1st Conference on Fairness, Accountability and Transparency. pp. 77–91. PMLR (2018)
3. Gardner, J., Brooks, C., Baker, R. : Evaluating the Fairness of Predictive Student Models Through Slicing Analysis. In : 9th International Conference on Learning Analytics & Knowledge. pp. 225–234. ACM, Tempe AZ USA (Mar 2019)
4. Hellas, A., Ihantola, P., Petersen, A., Ajanovski, V.V., Gutica, M., Hynninen, T., Knutas, A., Leinonen, J., Messom, C., Liao, S.N. : Predicting academic performance : A systematic literature review. In : 23rd Annual ACM Conference on Innovation and Technology in Computer Science Education. p. 175–199. ACM, New York, NY, USA (2018)
5. Hu, Q., Rangwala, H. : Towards Fair Educational Data Mining : A Case Study on Detecting At-risk Students. In : 13th International Conference on Educational Data Mining. p. 7 (2020)
6. Korkmaz, C., Correia, A.P. : A review of research on machine learning in educational technology. Educational Media International **56**(3), 250–267 (2019)
7. Kuzilek, J., Hlosta, M., Zdrahal, Z. : Open university learning analytics dataset. Sci Data 4 **170171** (2017)
8. Lee, H., Kizilcec, R.F. : Evaluation of Fairness Trade-offs in Predicting Student Success. arXiv :2007.00088 [cs] (Jun 2020)
9. Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., Galstyan, A. : A Survey on Bias and Fairness in Machine Learning. arXiv :1908.09635 [cs] (Jan 2022)
10. Verger, M. : Investiguer la notion d'équité algorithmique dans les environnements informatiques pour l'apprentissage humain. In : 9ièmes RJC EIAH 2022 : Environnements Informatiques pour l'Apprentissage Humain (2022)
11. Verger, M., Lallé, S., Bouchet, F., Luengo, V. : Is Your Model "MADD" ? A Novel Metric to Evaluate Algorithmic Fairness for Predictive Student Models. In : 16th International Conference on Educational Data Mining (2023)
12. Verma, S., Rubin, J.S. : Fairness definitions explained. 2018 IEEE/ACM International Workshop on Software Fairness (FairWare) pp. 1–7 (2018)

EIAH2023 : 11ème Conférence sur les Environnements Informatiques pour l'Apprentissage Humain

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

12-16 juin 2023 Brest (France)

Session de posters

Instrumentation de l'association de registres sémiotiques dans un assistant de preuve

Emmanuel Beffara[1], Martin Bodin[2], Nadine Mandran[1], and Rémi Molinier[3]

[1] Univ. Grenoble Alpes, CNRS, Grenoble INP, LIG, 38000 Grenoble, France
[2] Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LIG, 38000 Grenoble, France
[3] Univ. Grenoble Alpes, CNRS, Institut Fourier, F-38000 Grenoble, France

Résumé Les assistants de preuve sont des outils informatiques conçus pour accompagner l'écriture de démonstrations mathématiques formalisées et leur utilisation pour l'enseignement de la preuve fait l'objet de diverses expérimentations en université. Leur mode d'interaction se limite habituellement au registre textuel, à l'opposé de la pratique courante sur papier qui exploite souvent différents registres visuels. Ce travail propose la conception et l'évaluation d'un prototype d'interface qui permette d'associer des registres textuels et visuels pour améliorer la flexibilité de l'outil et faciliter l'apprentissage de la preuve en mathématiques.

Mots-clefs. enseignement des mathématiques · registres sémiotiques · instrumentation · assistant de preuve

NB : Une version avec figures se trouve sur `https://hal.science/hal-04096240`.

1 Introduction

La preuve mathématique est un objet d'un haut niveau d'abstraction dont l'acquisition est un enjeu d'enseignement majeur à l'entrée à l'université. La didactique met en évidence le rôle crucial de l'articulation entre les différents rôles qu'elle joue [8] : au delà de la validation des assertions (vérification, explication), la preuve a un rôle de communication (présentation systématique des résultats et transmission à une communauté) et elle intervient dans le processus de recherche (exploration des conséquences d'une hypothèse, élaboration de conjectures, découverte de résultats). À chaque rôle correspondent des savoir-faire différents, ce qui fait de la preuve un objet dont la compréhension et l'enseignement restent difficiles. De plus, au passage du secondaire à l'université correspond un saut dans le niveau d'exigence formel quant à la production de démonstrations, avec un passage de l'argumentation à du raisonnement essentiellement déductif [1].

Au vu de ces difficultés, il est naturel de vouloir instrumenter la preuve comme on instrumente d'autres objets fondamentaux des mathématiques. En effet, calculatrices, tableurs, traceurs de courbes, logiciels de géométrie ou de calcul formel sont présents en classe et intégrés aux usages. Des assistants de preuve, logiciels dédiés à l'élaboration interactive de démonstrations, ont été développés depuis la fin des années 1960, initialement pour la recherche. Ils visent principalement des experts, mais le besoin d'élargir le public visé a conduit au développements d'assistants (cf. Mizar, CalcCheck) dont la syntaxe est lisible par des mathématiciens qui ne connaissent pas l'outil.

Après des premières expériences dans les années 2000 [16,4], l'intégration d'assistants de preuve dans l'enseignement supérieur est expérimenté depuis quelques années dans différentes universités [10]. On observe généralement un meilleur engagement des étudiants dans la recherche de preuve avec l'outil que dans une modalité classique sur papier, ce qui s'explique notamment par la présence de rétroactions immédiates, et la facilitation du tâtonnement. Les assistants de preuve semblent favoriser l'utilisation correcte de la logique et du symbolisme mais peuvent aussi conduire les étudiants vers de la preuve très formelle, renforçant les aspects syntaxiques au détriment du sens. En effet, du fait de leur grande technicité, ces outils peuvent faire obstacle à l'apprentissage des notions visées, au risque que l'enseignement instrumenté par assistant de preuve devienne un enseignement de l'outil et de ses spécificités plutôt que de la preuve mathématique en tant que telle. Face à ce problème, plusieurs outils (cf. Edukera, D∃∀duction) ont été conçus en simplifiant l'interface proposée à l'étudiant, au prix d'une limitation des possibilités offertes en termes de démonstration.

Une autre difficulté est liée à l'écart entre la pratique courante de la preuve et le cadre formel imposé par les assistants. Dans la pratique experte, la preuve exploite souvent des changements de registre sémiotique [6] avec l'emploi de figures, schémas, diagrammes, etc. Chaque registre a ses règles et conventions propres qui en font un outil de compréhension et de communication et l'association de différents registres enrichit la compréhension en variant les représentations d'un même objet. Dans l'enseignement, l'articulation entre le schéma et la preuve textuelle est centrale, dès le collège avec la géométrie, puis au lycée avec par exemple l'emploi de courbes ou de tableaux, puis à l'université dans différentes branches des mathématiques. L'utilisation pertinente des différents registres fait donc partie des enjeux de l'enseignement des mathématiques.

Pourtant, en dehors de la géométrie où le registre visuel est intrinsèque à l'objet d'étude, les assistants de preuve sont généralement focalisés sur la production de démonstrations purement textuelles, sous forme de programmes (pour les outils experts) ou de rédactions idéalisées (pour les outils à visée pédagogique) et les registres visuels restent alors au brouillon et à la charge de l'apprenant, sans statut explicite au sein de la preuve. Dans l'optique d'une instrumentation de l'apprentissage de la preuve, il semble donc important de prendre en compte cette nécessaire association de différents registres.

2 Travaux connexes

En géométrie, où la dimension visuelle est inhérente au sujet, beaucoup de travaux ont été menés en EIAH et en didactique, depuis l'introduction de la géométrie dynamique [11] où la préoccupation pour la preuve a conduit à la réalisation de micro-mondes pour le raisonnement géométrique [12]. Des travaux plus récents ont formalisé la géométrie euclidienne dans des assistants de preuve et fourni des interfaces exploitant des outils existants de géométrie dynamique [7,13]. En dehors de la géométrie, peu de développements exploitent l'association de différents registres. Des expériences existent, comme ProofWeb [9] qui affiche la preuve en cours sous forme d'arbre de dérivation ou Alectryon [14] qui permet de représenter visuellement des objets manipulés par l'assistant de preuve Coq [17], ce qui permet d'incorporer des éléments graphiques dans l'affichage au lieu de notations lourdes ou inhabituelles.

Ces éléments viennent en support visuel, afin de mieux comprendre l'état interne de l'assistant de preuve : il est toujours nécessaire d'utiliser des commandes textuelles pour interagir. En cela, ces approches ne changent pas radicalement le registre utilisé pour l'interaction avec l'outil.

La communauté Coq s'est montrée intéressée par ces différents outils, même en mettant de côté leur intérêt pédagogique. En effet, ajouter un support graphique (même partiel) aide à mieux communiquer sur des résultats certifiés. De plus, certains domaines s'appuient naturellement sur des diagrammes, et ce support pourrait faciliter le développement des bibliothèques associées (notamment en théorie des catégories [3]).

3 Proposition de contribution

Notre premier objectif est de construire un prototype pour assister l'enseignement de la preuve en exploitant l'articulation entre plusieurs registres. Une attention particulière est portée sur l'objectif de faire de l'assistant de preuve un instrument d'apprentissage appropriable par l'étudiant, dans un mécanisme de genèse instrumentale [15]. Nous nous basons pour cela sur la Théorie des situations didactiques [5] et notamment le modèle cK¢ [2] qui vise à l'opérationnaliser : ce cadre permet d'identifier les caractéristiques de l'EIAH en tant que milieu et les variables didactiques qu'elles induisent (complexité des raisonnements nécessaires, possibilités d'interaction, degré d'automatisation, etc.) et par la suite permettra une modélisation des connaissances de l'apprenant par l'analyse des stratégies de résolution qu'il adpote. En conséquence, un second objectif est de valider la pertinence de ces modèles dans le contexte particulier des assistants de preuve.

Dans cette première version, nous concevons une implémentation avec Coq des tableaux de variation destinée à être manipulée par des étudiants de début d'université. Il s'agit d'une représentation visuelle d'un ensemble d'assertions logiques associées à une méthode de démonstration, l'agencement de l'information dans le tableau est porteur de sens et sa construction guide et structure le raisonnement. Par ailleurs, il fait partie de la pratique courante au sortir du lycée, ce qui permet de s'appuyer sur un savoir-faire existant chez les étudiants.

L'outil envisagé permettra de naviguer librement entre le registre visuel du tableau, où l'étudiant peut dialoguer avec l'assistant dans une interface graphique, et le registre textuel habituel de l'assistant. On peut alors envisager un spectre d'utilisations de l'outil, du purement graphique (avec restitution éventuelle d'une preuve écrite) au purement textuel (où l'outil produit une illustration graphique évoluant à chaque étape).

4 Méthode de conduite de la recherche

Nous conduisons nos recherches selon les principes du Design-Based Research [18] ainsi que les guides associés (citation d'auteur). Cette méthode propose, entre autres, de construire les connaissances et les outils associés de manière itérative en intégrant les acteurs du terrain. Ainsi, nous organiserons des focus-groups avec des étudiants pour évaluer le prototype et ensuite nous mènerons des études in situ en début de cycle à l'université. Nous nous interrogerons sur l'utilisabilité de l'outil, l'engagement des étudiants dans cette activité et leur utilisation de l'association des registres. Les

premiers résultats attendus sont une version du prototype et les avis des étudiants ayant participé au focus-group.

Références

1. Balacheff, N. : L'argumentation mathématique, précurseur problématique de la démonstration. In : XXVIe Colloque CORFEM. Strasbourg, France (Jun 2019)
2. Balacheff, N., Margolinas, C. : cK¢ Modèle de connaissances pour le calcul de situations didactiques. In : XII° école d'été de didactique des mathématiques. p. 1. La Pensée Sauvage éditions (Aug 2003)
3. Behr, N. : ANR CoREACT. https ://coreact.wiki/
4. Blanc, J., Giacometti, J., Hirschowitz, A., Pottier, L. : Proofs for freshmen with Coqweb. In : International Workshop on Proof Assistants and Types in Education (2007)
5. Brousseau, G. : Théorie des situations didactiques. La Pensée Sauvage, Grenoble (1998)
6. Duval, R. : Registres de représentation sémiotique et fonctionnement cognitif de la pensée. Annales de didactique et de sciences cognitives **5**, 37–65 (1993)
7. Guilhot, F. : Formalisation en Coq et visualisation d'un cours de géométrie pour le lycée. Techniques et sciences informatiques **24**(9), 1113–1138 (Nov 2005). https://doi.org/10.3166/tsi.24.1113-1138
8. Hanna, G. : Proof, Explanation and Exploration : An Overview. Educational Studies in Mathematics **44**(1/2), 5–23 (2000). https://doi.org/10.1023/A:1012737223465
9. Hendriks, M., Kaliszyk, C., van Raamsdonk, F., Wiedijk, F. : Teaching logic using a state-of-the-art proof assistant. Acta Didactica Napocensia **3**(2), 35–48 (2010)
10. Kerjean, M., Le Roux, F., Massot, P., Mayero, M., Mesnil, Z., Modeste, S., Narboux, J., Rousselin, P. : Utilisation des assistants de preuves pour l'enseignement en L1. Gazette de la Société Mathématique de France **174** (Oct 2022)
11. Laborde, C., Capponi, B. : Cabri-géomètre constituant d'un milieu pour l'apprentissage de la notion de figure géométrique. Recherches en didactique des mathématiques **14**(1.2), 1–15 (1994)
12. Luengo, V. : Analyse et prise en compte des contraintes didactiques et informatiques dans la conception et le développement du micromonde de preuve Cabri-Euclide. Sciences et Techniques Educatives **6**(2), 27 (1999)
13. Pham, T.M., Bertot, Y. : A Combination of a Dynamic Geometry Software With a Proof Assistant for Interactive Formal Proofs. Electronic Notes in Theoretical Computer Science **285**, 43–55 (Sep 2012). https://doi.org/10.1016/j.entcs.2012.06.005
14. Pit-Claudel, C. : Untangling mechanized proofs. In : Proceedings of the 13th ACM SIGPLAN International Conference on Software Language Engineering. pp. 155–174. ACM, Virtual USA (Nov 2020). https://doi.org/10.1145/3426425.3426940
15. Rabardel, P. : Les Hommes et Les Technologies ; Approche Cognitive Des Instruments Contemporains. Armand Colin (1995)
16. Raffalli, C., David, R. : Computer Assisted Teaching in Mathematics. In : Workshop on Thirty Five Years of Automath. Edinburgh (2002)
17. The Coq development team : Coq (1989)
18. Wang, F., Hannafin, M.J. : Design-based research and technology-enhanced learning environments. Educational Technology Research and Development **53**(4), 5–23 (Dec 2005). https://doi.org/10.1007/BF02504682

Intégration d'explications dans un référentiel de compétences associé à une plateforme d'apprentissage de l'informatique

Esther Félix[1], Franck Amadieu[2], Rémi Venant[3] et Julien Broisin[1]

[1] IRIT, Université de Toulouse, CNRS, Toulouse INP, UT3, Toulouse, France
{esther.felix, julien.broisin}@irit.fr
[2] CLLE - Cognition, Langues, Langage, Ergonomie
franck.amadieu@univ-tlse2.fr
[3] LIUM - Laboratoire d'Informatique de l'Université du Mans
remi.venant@univ-lemans.fr

Résumé. Cet article concerne le domaine de l'intelligence artificielle explicable appliquée à l'éducation, qui explore les explications destinées aux utilisateurs d'environnements informatiques pour l'apprentissage humain. Nous proposons d'introduire des explications dans un référentiel de compétences d'une plateforme d'apprentissage de l'informatique et nous étudions l'impact de certaines caractéristiques des apprenants sur leur perception et leur consultation des explications. Nos résultats montrent que la perception des étudiants de leur capacité à réussir est corrélée positivement avec le nombre d'explications ouvertes et que les étudiants les plus actifs sur la plateforme sont aussi ceux qui accèdent le plus aux explications.

Mots-clés : Explicabilité, Apprentissage de l'informatique, EIAH, Personnalisation, Intelligence Artificielle.

Abstract. This paper concerns the field of explainable artificial intelligence applied to education, which explores explanations for users in Technology-Enhanced Learning Environments. We propose to introduce explanations into a competency profile of a computer science learning platform and study the impact of some learners' characteristics on their perception and consultation of explanations. Our results show that students' perception of their ability to succeed is positively correlated with the number of opened explanations and that the most active students on the platform are also the ones who access explanations the most.

Keywords: Explainability, Computer Education, TEL, Personalization, Artificial Intelligence.

1 Introduction

Les Environnements Informatiques pour l'Apprentissage Humain (EIAH) sont souvent accompagnés de systèmes qui utilisent des méthodes d'intelligence artificielle pour interagir avec les apprenants ou les professeurs, par exemple en leur faisant des recommandations. Ces systèmes, peu transparents, peuvent bénéficier de l'ajout d'explications afin d'augmenter la confiance des utilisateurs en

améliorant leur compréhension des mécanismes sous-jacents à ces systèmes [2]. Pour être efficaces, les explications fournies doivent être adaptées au niveau d'expertise de l'utilisateur et plus généralement à des caractéristiques propres à l'utilisateur [7]. Nous proposons d'ajouter des explications à un référentiel de compétences intégré dans un environnement virtuel d'apprentissage de l'informatique. Une étude expérimentale menée avec des étudiants en informatique dans le supérieur nous permet de déterminer le lien entre certaines caractéristiques personnelles des étudiants et leur consultation des explications.

2 Travaux connexes

L'explicabilité des modèles d'intelligence artificielle appliqué à l'éducation est un domaine de recherche qui vise à améliorer la transparence, l'interprétabilité et la confiance envers les systèmes d'intelligence artificielle (IA) utilisés dans des contextes éducatifs [2]. Cela implique d'expliquer le fonctionnement interne de ces systèmes et leurs processus décisionnels aux utilisateurs finaux, tels que les étudiants, les enseignants ou encore les parents. Conati et al. ont par exemple proposé des conseils accompagnés d'explications dans un ITS (*Intelligent Tutoring System*) en se focalisant sur *pourquoi* une suggestion est utile et *comment* l'IA prend la décision de faire cette suggestion [3]. De plus, les explications peuvent être adaptées en fonction de différentes capacités, traits de caractère à long terme, états cognitifs et affectifs ou préférences des utilisateurs [4]. Des caractéristiques de ce type ont déjà été identifiées dans des domaines tels que la recommandation de musiques [5], mais n'ont à notre connaissance pas été clairement identifiées dans le domaine de l'éducation [4].

3 Méthodes

Outils utilisés L'expérimentation a été effectuée en s'appuyant sur la plateforme Lab4CE [1]. Ce système intègre différents outils et visualisations permettant la réalisation à distance de travaux pratiques d'informatique. Lab4CE offre à chaque apprenant un laboratoire virtuel personnel, et intègre des fonctionnalités permettant de collecter des données reflétant les différentes actions réalisées par chacun des apprenants. Lab4CE intègre également une approche par compétences [6] : un référentiel de compétences est associé au domaine informatique étudié, et le taux de maîtrise des apprenants pour chacun des items du référentiel est calculé automatiquement à partir des activités qu'ils réalisent sur la plateforme. L'algorithme responsable du calcul du taux de maîtrise de chacune des compétences analyse les commandes et scripts exécutés sur Lab4CE par l'apprenant. Plus précisément, chaque compétence est associée à une ou plusieurs expressions régulières et l'algorithme vérifie si des commandes ou des scripts écrits par les étudiants et exécutés sans erreur correspondent à ces expressions régulières. Ainsi, le taux de maîtrise d'une compétence augmente en fonction du nombre de correspondances trouvées avec les expressions régulières associées à cette compétence. Les explications portant sur les taux de maîtrise sont spécifiques à chaque compétence et s'obtiennent en cliquant sur les valeurs des taux

de maîtrise. Ces explications informent que le taux de maîtrise dépend de l'exécution correcte de commandes ou scripts liés à cette compétence. Dans le cas où le taux de maîtrise n'est pas nul, l'explication fournit également un extrait de code ayant contribué à faire monter le taux.

Expérimentation : données collectées et analyses Nous avons mené une expérimentation dans le contexte d'un module de programmation Shell pour lequel des séances de travaux pratiques hebdomadaires étaient organisées durant 5 semaines avec 100 étudiants d'IUT informatique. Deux groupes expérimentaux de 50 étudiants chacun ont été constitués aléatoirement avant la première séance. Les étudiants du premier groupe avaient accès aux explications portant sur la manière dont le taux de maîtrise était calculé, au contraire des étudiants du deuxième groupe (groupe contrôle). Nous avons ensuite collecté des données concernant toutes les interactions des étudiants avec le système. Nous avons également collecté les notes des étudiants à l'examen final du module, ainsi que leurs réponses à un questionnaire comprenant des items de Likert sur l'utilité perçue, la compréhension et la confiance des étudiants envers le référentiel de compétence. De plus, les trois caractéristiques personnelles des apprenants que nous avons pu mesurer avant le début du cours sont leur intérêt pour le thème du module étudié, leur perception de leur capacité à réussir le module, et leur niveau initial sur les connaissances théoriques du module. L'intérêt pour le thème du module et la perception de la capacité à réussir ont été mesurées avec plusieurs items de Likert. Le niveau initial est quant à lui représenté par la note au questionnaire de connaissances. L'analyse des données a ensuite été effectuée à l'aide des outils jamovi et Jupyter Notebook (lang. Python) avec les bibliothèques numpy et pandas. Des corrélations ont été effectuées entre le nombre d'ouvertures d'explications et, d'une part, les caractéristiques personnelles mesurées, et d'autre part les diverses données d'activité sur la plateforme, et les deux groupes expérimentaux ont été comparés sur leurs notes à l'examen final ainsi que sur leurs réponses aux questionnaires de perception du référentiel à l'aide d'analyses ANOVA.

4 Résultats

L'analyse de corrélations effectuée sur l'ensemble des étudiants ayant reçu des explications fait apparaître une corrélation positive significative entre la variable de perception de la capacité à réussir et le nombre d'ouvertures d'explications (r de Pearson $= 0.219$, $p < 0.05$). Cependant, on n'obtient pas de corrélation entre la variable d'intérêt pour le thème du cours ou la note au questionnaire de connaissances et les ouvertures d'explications. Nous avons également relevé des corrélations significativement positives entre le nombre de consultations d'explications et toutes les variables d'activité sur Lab4CE (nombre d'ouvertures d'un éditeur de code depuis le terminal de la plateforme, commandes exécutées et scripts exécutés). Enfin, les analyses ANOVA n'ont pas permis de relever de différences significatives entre les deux groupes concernant leurs résultats académiques ou leur perception du référentiel de compétences

(utilité perçue, compréhension et confiance).

5 Discussion et conclusion

Nous recommandons d'approfondir les recherches visant à personnaliser les explications dans les EIAH en évaluant et en prenant en compte la perception qu'ont les apprenants de leur capacité à réussir. Concernant les autres caractéristiques, ni l'intérêt pour le cours ni le niveau de connaissances avant le début du cours ne semblent avoir de lien significatif avec la consultation des explications. De plus, les étudiants qui consultent plus les explications sont ceux qui sont par ailleurs les plus actifs sur les autres aspects de l'environnement virtuel. Il semble donc que nos explications atteignent en premier lieu les apprenants impliqués et actifs dans l'utilisation de l'outil. Cela souligne le besoin d'explorer des variations sur la forme et le contenu des explications afin de mieux adapter les explications aux différents profils d'apprenants. On peut par exemple imaginer la possibilité de faire apparaître plus explicitement et régulièrement des signaux menant aux explications pour les étudiants qui ont peu de probabilité de les consulter.

Pour atteindre l'objectif de fournir des explications accessibles, utiles et attrayantes pour les étudiants de tous niveaux, il est nécessaire d'analyser plus en détail les besoins de différents groupes d'apprenants avec des méthodes qualitatives. De plus, les résultats obtenus soulignent le besoin d'examiner d'autres caractéristiques personnelles pouvant modifier l'intérêt porté aux explications, et de proposer des explications adaptées aux apprenants en fonction de ces caractéristiques. Une piste d'amélioration des explications fournies consiste à donner des retours aux apprenants sur des commandes n'ayant pas fonctionné.

Références

1. Broisin, J., Venant, R., Vidal, P. : Lab4ce : a remote laboratory for computer education. International Journal of Artificial Intelligence in Education (2017)
2. Conati, C., Porayska-Pomsta, K., Mavrikis, M. : AI in Education needs interpretable machine learning : Lessons from Open Learner Modelling. ArXiv (2018)
3. Conati, C., Barral, O., Putnam, V., Rieger, L. : Toward personalized XAI : A case study in intelligent tutoring systems. Artificial Intelligence **298**, 103503 (2021)
4. Khosravi, H., Shum, S.B., Chen, G., Conati, C., Tsai, Y.S., Kay, J., Knight, S., Martinez-Maldonado, R., Sadiq, S., Gašević, D. : Explainable Artificial Intelligence in education. Computers and Education : Artificial Intelligence **3**, 100074 (Jan 2022)
5. 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 (Mar 2019)
6. Sablayrolles, L., Guin, N., Lefevre, M., Broisin, J. : Conception et évaluation d'un processus de personnalisation fondé sur des référentiels de compétences. In : 10e Conférence sur les Environnements Informatiques pour l'Apprentissage Humain (2021)
7. Tomsett, R.J., Braines, D., Harborne, D., Preece, A., Chakraborty, S. : Interpretable to Whom ? A Role-based Model for Analyzing Interpretable Machine Learning Systems. ArXiv (Jun 2018)

RLG Model et RLG kit : un modèle et un kit de conception de *Role Learning Game*

Gaëlle Guigon[1,2], Mathieu Vermeulen[2], Mathieu Muratet[1], Thibault Carron[1]

[1] Sorbonne Université, CNRS, LIP6, F-75005 Paris, France
[2] IMT Nord Europe, IMT, Univ. Lille, CERI SN, F-59000 Lille, France
gaelle.guigon@imt-nord-europe.fr

Résumé Cet article présente un modèle de conception de *Role Learning Game* (RLG). Les RLG sont des jeux de rôles pédagogiques multijoueurs, multi-rôles. Ce type de jeu commence à se développer, mais les enseignants ont rarement les outils pour en concevoir eux-mêmes. Basé sur un modèle, un kit de conception de scénarios de RLG a été créé. Nous avons mené deux expérimentations qualitatives pour évaluer l'utilisabilité du kit. L'une des expérimentations a permis d'étudier le processus dans sa globalité, de la conception du RLG avec le kit jusqu'à sa mise en place avec des élèves. L'analyse thématique des sessions a permis de relever des pistes d'amélioration pour une extension du kit.

Mots-clés Jeux de rôles · Scénario · Enseignement · Multijoueur · Jeu d'apprentissage.

Abstract This article presents a design model for Role Learning Games (RLG). RLGs are multi-player, multi-role role-playing learning games. This type of game is developing, but teachers rarely have the tools to design one themselves. Based on a model, a RLG scenario design kit was created. We submitted the kit to two qualitative experiments to assess its usability. One of the experiments enables to study the process, from the RLG design phase with the kit to its implementation with students. The thematic analysis made it possible to identify areas for improvement for a kit expansion.

Keywords Role-playing game · Scenario · Education · Multiplayer · Learning game.

1 Introduction

Dans un monde en constante évolution, où chaque individu doit endosser des rôles différents tout au long de sa carrière avec une dimension sociale conséquente, l'importance de former des citoyens capables de s'adapter et de coopérer se fait de plus en plus sentir. Les jeux sérieux ont déjà montré leur intérêt pour appréhender des situations réelles, en particulier pour le travail en commun et

la co-construction des savoirs [2]. Il semble donc pertinent de se demander comment construire des jeux sérieux multijoueur multi-rôles et, qui plus est, comment favoriser leur construction par les enseignants qui préconisent leur usage. De nombreuses complexités pour concevoir les jeux existent [4]. À ces difficultés s'ajoute celle, non triviale, du multijoueur comme il est montré par Wendel [5], les *Learning Games* (LG) multijoueurs doivent répondre aux contraintes des jeux à joueur unique, mais également intégrer les défis liés au multijoueur et à la conception des LG. De surcroit, les méthodes pour aider à la conception de ce type de jeu semblent limités [3]. Enfin, la complexité s'accroît encore si on considère des rôles distincts interagissant simultanément au sein des LG multijoueurs. Notre objectif est donc de donner les clés aux enseignants pour qu'ils puissent concevoir plus facilement des LG multijoueurs de type jeux de rôles, *a fortiori* distincts. La problématique développée dans ce papier est de définir un modèle et un outil pour faciliter la conception de *Role Learning Games* (RLG), des jeux de rôles asymétriques à visée pédagogique. Pour répondre à cette question, nous avons mené deux expérimentations avec des enseignants de niveaux différents et de disciplines variées. Dans un premier temps, nous détaillerons notre contribution puis nous présenterons la méthode utilisée. Ensuite, nous aborderons les résultats et nous les analyserons avant de conclure.

2 Contribution

Nous proposons un modèle et sa réification sous forme de kit pour le rendre plus accessible à notre public cible : les enseignants de tous niveaux, ainsi qu'aux ingénieurs pédagogiques qui pourront les accompagner dans ce projet.

RLG Model : Nous avons construit un modèle (voir Fig. 1) représentant les

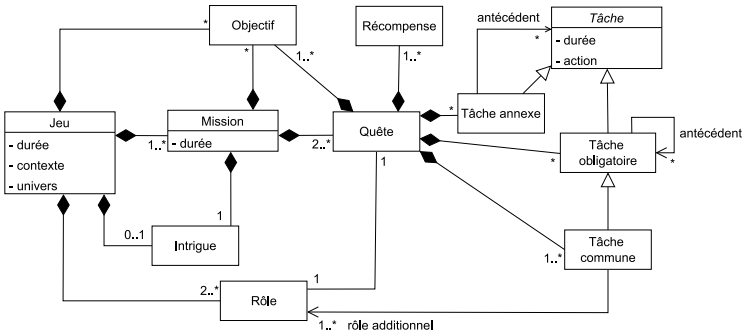


Fig. 1. Le RLG model (sous la forme d'un diagramme de classes UML) représente la configuration d'un scénario et l'organisation de ses missions et tâches.

principaux éléments à prendre en compte lors de la conception d'un RLG. Dans ce modèle, un jeu a une durée, un contexte et un univers. Ce jeu peut avoir un

ou plusieurs objectifs globaux, une intrigue générale, et implique nécessairement au moins deux rôles distincts dans une ou plusieurs missions. Chaque mission a une durée définie, une intrigue, et peut viser un ou plusieurs objectifs. Une mission est composée *a minima* de deux quêtes, une pour chaque rôle impliqué dans cette mission. Les quêtes, associées à un unique rôle, vont permettre d'atteindre une récompense et de répondre aux objectifs définis en réalisant des tâches. Chaque quête comporte au moins une tâche commune avec un autre rôle de sa mission puisque l'aspect coopératif est central dans ce type de jeu. Il peut y avoir d'autres tâches nécessaires pour accomplir la quête, mais également des tâches annexes permettant d'approfondir une tâche déjà réalisée. Le but de ces tâches annexes est de réguler les moments de synchronisation entre les joueurs lors des tâches communes (ajuster la durée de quêtes, occuper un joueur trop rapide, etc). Chaque tâche a une durée estimée ainsi qu'une action à réaliser (qui peut être de haut niveau). L'enseignant peut régler la granularité du jeu en fonction de sa complexité, et ainsi détailler plus ou moins les actions à réaliser. Enfin, une tâche (obligatoire ou annexe) peut avoir pour antécédent une autre tâche, une tâche obligatoire ne peut être antécédente qu'à une autre tâche obligatoire.

RLG Kit : Ce kit[3] se présente sous la forme de pièces de puzzle réinscriptibles à assembler dans un certain ordre. Un guide sous forme de livret définit la procédure à suivre et donc l'ordre dans lequel compléter les pièces pour créer un scénario de RLG. Le dernier élément du kit est une fiche exemple de scénarios réalisés avec ce kit.

3 Méthode

Ces deux expérimentations ont été menées avec des enseignants de niveaux différents, pour des disciplines distinctes. À la fin des expérimentations, les testeurs ont rempli le questionnaire SUS [1] pour évaluer l'utilisabilité du kit. Pour vérifier si cet outil était compréhensible et suffisant, nous avons mené les sessions avec des non-experts de la création de LG. Le but était de réussir à concevoir un scénario de RLG avec l'outil. La première expérimentation a été réalisée avec un enseignant de lycée pour un cours de français. La deuxième expérimentation a réuni sept enseignants de collège dans des disciplines diverses.

Les expérimentations avaient pour but de répondre à ces questions : 1) Les utilisateurs comprennent-ils comment utiliser chaque élément du kit ? 2) Est-ce que les testeurs réussissent à créer un scénario de RLG avec ce kit ? 3) Y a-t-il des améliorations à apporter au modèle ?

Pour la première expérimentation, la prise en main du kit a été très rapide, le kit s'est révélé être d'une grande utilité pour ordonner les tâches entre les rôles, pour repérer le nombre d'unités de temps entre les tâches communes, et pour comparer les tâches à effectuer (notamment entre rôles). L'enseignant a abouti à un scénario complet au bout des quatre heures d'expérimentation [4]. Le jeu

[3] Kit RLG : https://bit.ly/43UYqVa, accédé le 21/04/2023

[4] Scénario de la 1re expérimentation : https://bit.ly/3H1edbz, accédé le 12/01/2023

a été mis en oeuvre avec 16 élèves. Durant la seconde expérimentation de trois heures, les enseignants ont rapidement compris le fonctionnement des tuiles. Le jeu (en cours de réalisation), a pour but de faire réviser le brevet à des élèves de troisième.

Concernant les résultats[5] du formulaire SUS de Brooke, la moyenne des résultats est de 80/100 pour les huit enseignants. Ce résultat est donc plutôt satisfaisant, même si nous envisageons quelques améliorations.

4 Retour d'expérience et conclusion

Suite aux résultats des expérimentations, quelques nouvelles tuiles vont être ajoutées dans une extension pour améliorer le kit, mais aucun élément primordial n'était manquant. Voici quelques éléments de réponse aux questions vues en 3 : 1) Les expérimentations menées tendent à montrer que les utilisateurs comprennent rapidement comment utiliser le kit. 2) La première expérimentation a montré que l'enseignant avait réussi à créer un scénario de RLG avec ce kit. 3) Suite à ces expérimentations, il n'y a pas d'élément primordial manquant au modèle, cependant, comme cité précédemment, plusieurs améliorations seront implémentées dans le modèle et dans le kit.

Pour conclure, nous proposons un modèle de conception de *Role Learning Games*. Ce dernier a été réifié sous la forme d'un kit de conception. Ce kit a été expérimenté de manière qualitative pour valider le modèle. Les résultats sont très satisfaisants et un des RLG scénarisés a pu être joué en cours avec des élèves. Le jeu s'est déroulé comme prévu et les élèves comme l'enseignant sont ravis et souhaitent reproduire l'expérience dès que possible. Les testeurs ont rapidement pris en main le kit et ont vite compris son fonctionnement. Ces expérimentations ont permis de constater que l'outil sous forme de puzzle réinscriptible semble aider les enseignants dans la conception de RLG et qu'il était possible de réaliser des RLG avec le kit. Ces expérimentations nous ont également donné des pistes d'amélioration pour guider davantage les enseignants novices dans la conception de jeux.

Références

1. Brooke, J. : SUS a 'Quick and Dirty' Usability Scale. In : Usability Evaluation in Industry, pp. 189–194. Taylor & Francis (Jan 1996)
2. Leclerc, G., Bommel, P., Motisi, N., Vezy, R., Treminio, E., Avelino, J. : Coffee leaf rust (Hemeleia vastatrix) risk management in Central America : contribution of remote interactive simulations. Agronomie, Environnement & Sociétés **11**(2) (Dec 2021)
3. Mariais, C. : Modèles pour la conception de Learning Role-Playing Games en formation professionnelle (Apr 2012)
4. Vermeulen, M. : Une approche meta-design des learning games pour développer leur usage. phdthesis, Sorbonne Université (Jun 2018)
5. Wendel, V., Gutjahr, M., Göbel, S., Steinmetz, R. : Designing collaborative multiplayer serious games. Educ Inf Technol **18**(2), 287–308 (Jun 2013)

[5] Résultats du SUS : https://bit.ly/3H1rrVO, accédé le 12/01/2023

Perspectives d'analyse de commentaires métacognitifs pour la conception d'un EIAH

Christian Hoffmann, Julien Douady et Nadine Mandran

Univ. Grenoble Alpes, CNRS, Grenoble INP, LIG
prenom.nom@univ-grenoble-alpes.fr

Résumé. Le travail présenté s'inscrit dans le cadre de la conception d'un outil informatique pour assister les enseignants dans le développement des aptitudes métacognitives de leurs élèves. Dans notre recherche, il est demandé aux élèves de formuler des commentaires métacognitifs suite à une évaluation formative. Le poster résumera les résultats de trois différentes méthodes pour analyser ces commentaires, ainsi que les perspectives associées en vue d'une analyse automatisée.

Mots-clés : métacognition, évaluation formative, degrés de certitude, analyse qualitative, EIAH.

Abstract. The work presented is part of the design of a learning technology to assist teachers in the development of their students' metacognitive skills. In our research, students are asked to provide metacognitive feedback following a formative assessment. The poster will summarize the results of three different methods to analyze these comments, as well as future prospects for automated analysis.

Keywords: metacognition, formative assessment, degrees of certainty, qualitative analysis, TEL systems.

1 Introduction

Selon la définition originale de Flavell, la métacognition est « la connaissance et la régulation des activités cognitives de l'individu dans les processus d'apprentissage » [1]. Dès les années 1990, la méta-analyse de Wang *et al.* [2] a révélé que la métacognition est un facteur prédictif puissant de l'apprentissage. De plus, selon la perspective socio-cognitive [3], la métacognition est également une composante essentielle dans l'autorégulation des apprenants ; il est donc important pour eux de développer de telles aptitudes. Si des apprenants le font spontanément en s'inspirant de leurs parents, professeurs ou pairs, d'autres ne le font pas et il devient alors nécessaire de l'enseigner [4]. L'objectif final de l'étude exploratoire présentée ici est de proposer un outil aux enseignants du supérieur et du secondaire pour leur faciliter cette tâche.

2 Problématique et Questions de Recherche

Nous exploitons dans cette communication les commentaires métacognitifs formulés par des élèves suite à une évaluation formative basée sur des Questions à Choix Multiples (QCM) avec des degrés de certitude (DC) [5]. En associant des DC aux QCM, le feedback renvoyé aux apprenants est plus riche et plus précis [6], *cf.* le tableau 1 pour un exemple, et stimule l'activité métacognitive.

Table 1. Exemple de feedback suite à une QCM en combinant justesse et degré de certitude

		Réponse	
		correcte	**incorrecte**
Certitude	**forte**	Connaissance solide => élève bien informé	Erreur insoupçonnée => élève mal-informé
	faible	Connaissance fragile => élève chanceux	Erreur présumée => élève non-informé

Le défi des enseignants consiste à amener les apprenants à exploiter le feedback fourni. Un moyen qui semble probant est d'inciter les élèves à écrire des commentaires métacognitifs au moment de la consultation du feedback. Nous développons un plug-in Moodle dans ce but. En parallèle, nous avons mené une étude préliminaire où nous avons récolté de tels commentaires *via* des questionnaires sur papier.

À ce stade, la question saillante est : Comment aider les enseignants dans l'analyse des commentaires métacognitifs, tout en conservant autant que possible l'avantage des QCM, à savoir une exploitation automatique des résultats ? Autrement dit, est-il possible d'automatiser la catégorisation et l'évaluation de la qualité des commentaires, au moins partiellement, dans un Environnement Informatique pour l'Apprentissage Humain (EIAH) ?

Dans cette communication, nous explorons trois méthodes d'analyse de données qualitatives : i) une analyse thématique, ii) une analyse automatique par logiciel et iii) une analyse en utilisant une catégorisation préexistante, basée sur les théories de l'autorégulation. Nous adressons les deux questions de recherche suivantes :

1) Le croisement des trois approches permet-il de dégager une catégorisation pertinente qui pourrait assister l'enseignant dans l'exploitation des commentaires ?
2) Quelles pistes se dessinent pour mettre en place une analyse (semi-)automatisée sur les plateformes d'apprentissage afin de fournir des rétroactions aux apprenants ?

3 Méthode de Recherche et Données

Puisque nous devons impliquer les enseignants et les élèves en situation réelle, nous mobilisons le Design Based Research [7], ses principes et son outillage par les guides de THEDRE [8]. Suivant ces principes, nous avons conduit différentes observations dans des lycées où nous avons coconstruit avec les enseignants les QCM. Les élèves répondaient aux QCM et associaient un degré de certitude à leurs réponses. A l'issue

de quoi, ils commentaient par écrit, sur un questionnaire papier, le feedback reçu pour leurs résultats globaux. Certains élèves ont été invités à commenter également leurs réponses à des questions individuelles avec deux types de feedback « Erreur insoupçonnée » ou « Connaissance fragile », *cf.* tableau 1. Nous avons recueilli deux jeux de données auprès des mêmes lycéens, à deux moments différents de l'année scolaire 2019/2020, *cf.* tableau 2.

Table 2. Vue d'ensemble des collectes de données

	Observation 1	**Observation 2**
Effectifs	7 classes / 178 élèves	4 classes / 105 élèves
Période	oct. à déc. 2019	fév. 2020
Nb de commentaires globaux exploitables	178	103
Nb d'élèves qui ont formulé au moins un commentaire sur des quest. indiv.	76	49

4 Analyse des Données et Résultats

Les commentaires produits par les élèves sont nombreux et de différentes natures. Certains se réfèrent aux apprentissages disciplinaires, d'autres aux stratégies d'apprentissage employées, etc. Voici un exemple : « *J'ai eu 30/40, je pensais être sur des résultats pour lesquels j'ai eu "erreur insoupçonnée". Je pense qu'il faut que je sois plus prudent. Je devrais relire mes cours de l'année précédente et du début d'année.* »

Afin de dégager une catégorisation pertinente, nous prévoyons une triangulation de trois approches d'analyses de données qualitatives :

1. **Analyse thématique**. Dans cette méthode, l'analyste fait appel à des dénominations, appelées « thèmes ». Il s'agit de répondre progressivement à la question « qu'y a-t-il de fondamental dans ce propos, dans ce texte, de quoi y traite-t-on ? » [9]. L'analyse thématique ne nécessite pas de modèle théorique *a priori* pour conduire l'analyse. Il s'agit ensuite d'élaborer des synthèses sur la base des thèmes, illustrées par les verbatims. Nous mobiliserons cette méthode pour décrire et identifier des régularités dans les commentaires métacognitifs des élèves.
2. **Analyse textuelle automatique**. Cette méthode inclut trois niveaux : l'analyse lexicale pour étudier le vocabulaire, l'analyse syntaxique pour étudier les relations de dépendance entre les mots, et ensuite l'analyse sémantique pour étudier le sens du discours. Nous mobiliserons les travaux d'analyse textuelle de M. Reinert [10] implémentés sur le package Rainette du logiciel R [11].
3. **Analyse avec des thèmes prédéfinis**. L'approche consiste à définir *a priori* les thèmes utilisés pour la catégorisation des commentaires, en nous basant sur un modèle de l'autorégulation des apprentissages [12]. Nous avons choisi le modèle de Winne et Hadwin [13] car il se caractérise par une perspective métacognitive marquée, et il a été utilisé dans d'autres études visant à implémenter des EIAH.

Basé sur des analyses préliminaires, on constate que rares sont les élèves qui se limitent à décrire uniquement leurs résultats. La grande majorité des apprenants aborde également des thèmes qui se classent dans une ou plusieurs des catégories suivantes :

a) croyances, dispositions, styles individuels ; b) facteurs de motivation ; c) connaissances disciplinaires ; d) connaissance de la tâche ; e) connaissances des stratégies d'apprentissage. Ici il s'agit de la catégorisation issue de l'approche iii). La communication présentera les résultats des trois approches et les premières conclusions sur l'apport de leur croisement pour mettre en place une catégorisation automatisée.

Références

1. Flavell, J.H.: Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. Am. Psychol. 34, 906□911 (1979).
2. Wang, M.C., Haertel, G.D., Walberg, H.J.: What Influences Learning? A Content Analysis of Review Literature. J. Educ. Res. 84, 30□43 (1990).
3. Zimmerman, B.J.: Self-regulation involves more than metacognition: A social cognitive perspective. Educ. Psychol. 30, 217□221 (1995).
4. Veenman, M.V.J., Van Hout-Wolters, B.H.A.M., Afflerbach, P.: Metacognition and learning: conceptual and methodological considerations. Metacognition Learn. 1, 3□14 (2006).
5. Butler, A.C., Karpicke, J.D., Roediger, H.L.: Correcting a metacognitive error: Feedback increases retention of low-confidence correct responses. J. Exp. Psychol. Learn. Mem. Cogn. 34, 918□928 (2008).
6. Curtis, D.A., Lind, S.L., Boscardin, C.K., Dellinges, M.: Does student confidence on multiple-choice question assessments provide useful information? Med. Educ. 47, 578□584 (2013).
7. Wang, F., Hannafin, M.: Design-based research and technology-enhanced learning environments. Educ. Technol. Res. Dev. 53, 5□23 (2005).
8. Mandran, N., Vermeulen, M., Prior, E.: THEDRE's Framework: Empowering PhD Candidates to Efficiently Implement Design-Based Research. Educ. Inf. Technol. (2022).
9. Paillé, P., Mucchielli, A.: L'analyse qualitative en sciences humaines et sociales. Armand Colin, Malakoff (2021).
10. Reinert, M.: Postures énonciatives et mondes lexicaux stabilisés en analyse statistique de discours. Lang. Société. 121□122, 189 (2007).
11. Barnier, J.: The Reinert Method for Textual Data Clustering, https://juba.github.io/rainette/
12. Panadero, E.: A Review of Self-regulated Learning: Six Models and Four Directions for Research. Front. Psychol. 8, 422 (2017).
13. Greene, J.A., Azevedo, R.: A Theoretical Review of Winne and Hadwin's Model of Self-Regulated Learning: New Perspectives and Directions. Rev. Educ. Res. 77, 334□372 (2007).

De l'évaluation classique à l'évaluation adaptative

Adel Ihichr, Omar Oustous, Younès El Bouzekri El Idrissi, Ayoub Ait Lahcen

Laboratoire des Sciences de l'Ingénieur, ENSA, Université Ibn Tofail, Kénitra, Maroc
adel.ihichr@uit.ac.ma,omar.oustous@uit.ac.ma,
y.elbouzekri@uit.ac.ma,ayoub.aitlahcen@uit.ac.ma

Résumé : COVID19 a changé la doctrine de l'évaluation des connaissances du point de vue technologique. Cette évaluation constitue la pièce maîtresse d'un apprentissage adaptatif, car elle offre un moyen pour déterminer efficacement et effectivement les connaissances actuelles d'un apprenant pour identifier ses faiblesses et ses lacunes. Cet article vise à donner à tout chercheur en EIAH une vision holistique de l'utilisation de l'évaluation adaptative dans l'apprentissage adaptatif en passant en revue les plus célèbres techniques et algorithmes utilisées dans cette évaluation, ainsi que ses principales caractéristiques et avantages.

Mots-clés : évaluation, test adaptatif, apprentissage, EIAH, algorithmes

Abstract. COVID19 has changed the doctrine of assessment from a technological standpoint. This assessment is the centerpiece of adaptive learning, it provides a tool to determine a learner's current knowledge and identify weaknesses and gaps effectively and efficiently. This paper aims to give any LMS researcher a holistic view of the use of adaptive assessment in adaptive learning by reviewing famous techniques and algorithms used in, as well as its main features and benefits.

Keywords: assessment, adaptive test, learning, LMS, algorithms

1 Introduction

Ces dernières années sont marquées par le développement d'un nombre important des environnements informatiques pour l'apprentissage humain (EIAH) pour surmonter les nouvelles circonstances imposées par la pandémie [1]. Par surcroît, l'évaluation et le feedback sont les piliers d'un EIAH intelligent [2]. Ils offrent la possibilité de personnaliser un processus d'apprentissage à un apprenant afin de s'adapter à différentes situations et circonstances [3], en individualisant les composantes du processus d'apprentissage : contenu, interface, évaluation [4].

Dans cet article, nous présentons en section 2 le rôle joué par l'évaluation dans l'apprentissage adaptatif. La section 3 met le focus sur l'évaluation adaptative ; dans la section 4, nous exposons les algorithmes et les techniques les plus utilisées dans l'évaluation adaptative avant de conclure en ouvrant les perspectives de développement de l'évaluation adaptative pour améliorer la qualité de l'éducation via les EIAH.

2 Le rôle de l'évaluation dans un apprentissage adaptatif

L'évaluation de l'apprenant est un facteur crucial pour le succès du processus d'apprentissage en ligne [5]. Dans ce sens, les auteurs de [6] ont défini les principaux résultats d'apprentissage évalués par un système d'apprentissage, à savoir les résultats d'apprentissage cognitifs, comportementaux et affectifs, et ils ont prouvé que ces trois aspects sont corrélés. Par surcroît, une évaluation peut être globalement divisée en deux types, premièrement, l'évaluation sommative ou l'évaluation de l'apprentissage, qui se produit à la fin du cycle d'apprentissage pour mesurer la maîtrise d'un module, deuxièmement, l'évaluation formative ou l'évaluation pour l'apprentissage qui se produit tout au long du cycle d'apprentissage pour affiner la trajectoire [7]. En général, l'évaluation sommative est dominante par rapport à l'évaluation formative [8], mais le dernier type a montré plus de potentiel dans l'apprentissage adaptatif [9].

3 De l'évaluation classique à l'évaluation adaptative

A l'instar de l'apprentissage, l'adaptabilité est également intégrée à l'évaluation elle-même. Contrairement aux tests conventionnels, une évaluation adaptative fournit un test court et personnalisé, ses items changent en fonction de la réponse [4]. Les psychomotriciens ont développé les tests adaptatifs informatisés (CAT) pour évaluer les étudiants par ordinateur en adaptant les tests à leurs profils [10], sachant qu'un système basé sur le CAT peut réduire la longueur d'un test d'au moins 60% [11]. Les auteurs de [12] et [13] ont identifié trois types de tests utilisés dans l'évaluation adaptative : le test pyramidal, permettant d'évaluer les apprenants sans passer un test préliminaire, ils passent donc le même test de difficulté moyen, puis la tâche suivante est liée à la réponse précédente; le test à niveaux variables qui programme la première tâche avec un niveau de difficulté défini par le tuteur, chaque niveau comporte un item, et la difficulté de l'item suivant dépend de la réponse précédente; le dernier est le test adaptatif stratifié dans lequel plusieurs strates de difficultés sont définis, chacune regroupe des items du test qui ont approximativement la même difficulté moyenne, ainsi l'item suivant est sélectionné dans la strate supérieure si la réponse précédente est correcte, sinon le système suggère un item de la strate inférieure. D'autre part, l'évaluation adaptative montre plusieurs avantages : dynamiser et individualiser le processus d'évaluation [4], réduire le nombre de questions dans un test [4], augmenter la motivation de l'apprenant en suggérant des questions plus faciles [12], favoriser l'autonomie [12], fournir des statistiques plus détaillées pour affiner la trajectoire d'apprentissage [4], réduire la tricherie car le test est adapté à chaque apprenant [14].

4 L'évaluation adaptative : Algorithmes et techniques

Dans cette section, nous allons dresser les plus célèbres algorithmes et techniques utilisées dans la mise en œuvre d'une évaluation adaptative :

Théories : IRT [15] qui tente de relier des actions observables des réponses à des caractéristiques non observables et KST [16] qui reconnaît les compétences qui sont réalisables sans maîtriser d'autres compétences.

Techniques basées sur les réseaux bayésiens RB : Les auteurs de [17] ont utilisé les RB pour fournir une évaluation adaptative de plusieurs caractéristiques dans un test unique. De plus, il y a l'approche BKT [18] qui tente de modéliser la compétence de l'apprenant en calculant la probabilité de maîtriser une compétence en se basant sur quatre probabilités : devinement, glissement, probabilité que la compétence soit déjà maîtrisée, et probabilité que la compétence soit apprise.
Techniques basées sur le modèle logistique : La première est la PFA [19] qui est principalement sensible au rapport relatif entre des réponses correctes et des réponses incorrectes dans une évaluation, ce qui permet d'affiner l'adaptabilité de l'évaluation, la deuxième est la KTM [20] qui est apte à modéliser la maîtrise des connaissances de l'étudiant en utilisant un ensemble clairsemé des poids pour toutes les caractéristiques impliquées dans l'évaluation.
Techniques basées sur l'intelligence artificielle : La première est la DKT [21] qui permet de suggérer des évaluations plus adaptées aux besoins individuels et de sauter ou de retarder les questions qui s'annoncent trop faciles ou trop difficiles, et la deuxième est Fuzzy Logic [22] qui peut être utiliser pour fournir une évaluation adaptative, car elle peut traiter avec succès l'incertitude et la subjectivité humaine.
D'autre part, certains chercheurs tels que [23] ont préconisé qu'il y a peu de preuves pour soutenir l'idée que l'adaptation aux styles d'apprentissage individuels, à l'instar de FSLSM [13], améliore réellement les résultats de l'apprentissage.

5 Conclusion

Cet article est une investigation primitive de l'évaluation adaptative qui vise à nous fournir un aperçu sur la façon dont cette évaluation fonctionne, ses avantages, ses types et les différentes techniques utilisées dans sa mise en œuvre. Dans l'ensemble, cet article suggère que l'évaluation adaptative reste un champ prometteur pour mesurer les connaissances des élèves au bon endroit, au bon moment et sous la bonne forme ; cependant, nous constatons que les évaluations basées sur le jeu (GBDA) est l'une des techniques émergentes pour générer des résultats comportementaux authentiques et précis en menant une évaluation formative mais furtive.

Remerciements

Ce travail a été soutenu par le MESRSI, l'ADD et le CNRST du Maroc dans le cadre du programme Al-Khawarizmi, projet n°451/2020 (Smart Learning).

Références

1. UNESCO: Reopening and reimagining universities, survey on higher education through the UNESCO National Commissions, https://unesdoc.unesco.org/ark:/48223/pf0000378174, last accessed 2023/04/20.
2. Admiraal, W., Huisman, B., Pilli, O.: Assessment in massive open online courses. Electronic Journal of E-learning. 13, pp207□216-pp207□216 (2015).

3. Li, H., Cui, W., Xu, Z., Zhu, Z., Feng, M.: Yixue Adaptive Learning System and Its Promise on Improving Student Learning. In: CSEDU (2). pp. 45–52 (2018).
4. Morze, N., Varchenko-Trotsenko, L., Terletska, T., Smyrnova-Trybulska, E.: Implementation of adaptive learning at higher education institutions by means of Moodle LMS. J. Phys.: Conf. Ser. 1840, 012062 (2021).
5. Chrysafiadi, K., Troussas, C., Virvou, M.: Combination of fuzzy and cognitive theories for adaptive e-assessment. Expert Systems with Applications. 161, 113614 (2020).
6. Wei, X.: Assessment of cognitive, behavioral, and affective learning outcomes in massive open online courses: A systematic literature review. Computers & Education. 24 (2021).
7. Nikou, S.A., Economides, A.A.: Mobile-based assessment: A literature review of publications in major referred journals from 2009 to 2018. Computers & Education. 125, 101–119 (2018).
8. Shute, V.J., Rahimi, S.: Review of computer-based assessment for learning in elementary and secondary education: Computer-based assessment for learning. Journal of Computer Assisted Learning. 33, 1–19 (2017).
9. Csapó, B., Molnár, G.: Online Diagnostic Assessment in Support of Personalized Teaching and Learning: The eDia System. Front. Psychol. 10, 1522 (2019).
10. Oppl, S., Reisinger, F., Eckmaier, A., Helm, C.: A flexible online platform for computerized adaptive testing. Int J Educ Technol High Educ. 14, 2 (2017).
11. Lazarinis, F., Green, S., Pearson, E.: Creating personalized assessments based on learner knowledge and objectives in a hypermedia Web testing application. Computers & Education. 55, 1732–1743 (2010).
12. Lendyuk, T., Rippa, S., Sachenko, S.: Simulation of computer adaptive learning and improved algorithm of pyramidal testing. In: 2013 IEEE 7th International Conference on Intelligent Data Acquisition and Advanced Computing Systems (IDAACS). pp. 764–769. IEEE, Berlin, Germany (2013).
13. Al-Rajhi, L., Salama, R., Gamalel-Din, S.: Personalized Intelligent Assessment Model for Measuring Initial Students Abilities. In: Proceedings of the 2014 Workshop on Interaction Design in Educational Environments - IDEE '14. pp. 41–48. ACM Press, Albacete, Spain (2014).
14. Kosinski, M., Rust, J.: The development of Concerto: An open-sourceonline adaptive testing platform. Paper presenetd at the International Association for Computerized Adaptive Testing, Pacific Grove, CA. (2011).
15. Conejo, R., Guzmán, E., Pérez-De-La-Cruz, J.-L., Millán, E.: Introducing adaptive assistance in adaptive testing. 3 (2005).
16. Falmagne, J.-C., Albert, D., Doble, C., Eppstein, D., Hu, X. eds: Knowledge Spaces. Springer Berlin Heidelberg, Berlin, Heidelberg (2013).
17. Collins, J.A., Greer, J.E., Huang, S.X.: Adaptive assessment using granularity hierarchies and bayesian nets. In: Frasson, C., Gauthier, G., and Lesgold, A. (eds.) Intelligent Tutoring Systems. pp. 569–577. Springer Berlin Heidelberg, Berlin, Heidelberg (1996).
18. Corbett, A.T., Anderson, J.R.: Knowledge tracing: Modeling the acquisition of procedural knowledge. User Model User-Adap Inter. 4, 253–278 (1995).
19. Pavlik Jr, P.I., Cen, H., Koedinger, K.R.: Performance Factors Analysis--A New Alternative to Knowledge Tracing. Online Submission. (2009).
20. Vie, J.-J., Kashima, H.: Knowledge Tracing Machines: Factorization Machines for Knowledge Tracing. AAAI. 33, 750–757 (2019).
21. 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).
22. Jeremić, Z., Jovanović, J., Gašević, D.: Student modeling and assessment in intelligent tutoring of software patterns. Expert Systems with Applications. 39, 210–222 (2012).
23. Abyaa, A., Khalidi Idrissi, M., Bennani, S.: Learner modelling: systematic review of the literature from the last 5 years. Education Tech Research Dev. 67, 1105–1143 (2019).

Environnement virtuel de simulation dans la formation en Odontologie – Présentation d'un programme de recherche et des résultats d'une étude préliminaire

Valériane Loison[1], Fabrice Pirolli[1], Serena Lopez[2], Raphaëlle Crétin-Pirolli[1]

[1] Le Mans Université, CREN, F-53000 Laval, France
prenom.nom@univ-lemans.fr
[2] Nantes Université, CREN, F-44000 Nantes, France
serena.lopez@univ-nantes.fr

Résumé. La formation en Odontologie intègre, comme dans d'autres domaines de la santé, des environnements de simulation destinés à l'apprentissage des gestes techniques. Cet article présente les résultats préliminaires d'une étude menée au sein du programme de recherche EVAGO (Environnement Virtuel pour l'Apprentissage du Geste en Odontologie).

Mots-clés : Odontologie, simulation, haptique, réalité virtuelleformation

Abstract. The training in Dentistry integrates, as in other fields of health, simulation environments intended for learning technical gestures. This article presents the preliminary results of a survey conducted within the EVAGO research program (Virtual Environment for Learning to Gesture in Dentistry).

Keywords: Odontology, simulation, haptics, virtual reality, training

1 Introduction

Selon le Larousse, l'Odontologie concerne « l'étude des dents, de leurs maladies et du traitement de celles-ci ». La formation des chirurgiens-dentistes est dotée d'enseignements théoriques mais aussi pratiques qui permettent aux étudiants d'apprendre les gestes techniques de la dentisterie. Cette formation pratique s'appuie notamment sur l'utilisation de la simulation, qui répond au principe fondamental : « jamais la première fois sur le patient ». Le projet de recherche EVAGO (Environnement Virtuel pour l'Apprentissage du Geste en Odontologie) vise à étudier quels sont les usages de la réalité virtuelle pour l'apprentissage des gestes techniques et comprendre comment doit être développé et introduit ce nouveau type de simulateur dans la formation des étudiants en chirurgie dentaire. Il n'y a pas d'études comparables du point de vue de sa complexité et de sa transdisciplinarité. Nous présenterons dans un premier temps le programme de recherche EVAGO, ainsi que le simulateur étudié et la méthodologie de

recueil de données réalisée dans le cadre d'un premier protocole ; puis dans un second temps les résultats d'une étude quantitative préliminaire menée auprès des étudiants qui ont pu tester ce nouvel outil.

2 Programme de recherche EVAGO

Présentation du programme de recherche EVAGO. Quatre équipes de recherche et un partenaire industriel forment le Consortium EVAGO. Il implique des chercheurs et enseignants-chercheurs dans les domaines de l'Odontologie, de l'Informatique et des Sciences de l'Information et la Communication. Les objectifs du projet portent sur l'apport de nouvelles fonctionnalités au simulateur haptique dentaire Virteasy Dental® notamment l'évaluation automatique des gestes techniques. Au-delà de l'enrichissement sur le plan fonctionnel, la recherche doit permettre le développement de nouveaux scénarios et ressources pédagogiques afin de l'intégrer au mieux dans le cursus d'apprentissage des futurs chirurgiens-dentistes dans un cycle d'itération continue tout au long du projet. Plus largement, il s'agit d'un questionnement sur l'usage d'un environnement virtuel de simulation dans la formation en Odontologie. Basque et Doré (1998, p.5) [1] définissent un environnement comme *« le lieu qui abrite un système avec ses sous-systèmes, ce lieu pouvant être réel ou virtuel »*. Les Environnements Numériques d'Apprentissage (ENA) font l'objet de recherches visant à *« concevoir, développer et évaluer des environnements meilleurs, plus performants, plus adaptés, exploitant les dernières technologies développées »* selon Charlier (2019, p.51) [2]. Cette dimension du projet est conduite grâce à des analyses de pratiques et d'observations en contexte de formation réalisés par les différents acteurs du projet.

Présentation du simulateur Virteasy Dental®. Toutes les facultés de chirurgie-dentaire françaises utilisent le simulateur non numérique dit « fantôme ». Celui-ci prend la forme d'un modèle de mannequin avec un buste et une tête dans laquelle nous pouvons insérer des modèles de mâchoires. Il est utilisé pour la pratique préclinique des étudiants essentiellement en deuxième et troisième année de la formation en Odontologie. Cependant d'autres types de simulateurs voient le jour notamment équipés de la réalité virtuelle. *Virteasy Dental*® est un simulateur de chirurgie dentaire conçu par la société HRV. Il est destiné aux étudiants et à leurs enseignants pour favoriser l'apprentissage des gestes techniques de soins dentaires. Ce simulateur est un environnement composé de plusieurs dispositifs comprenant une tour d'ordinateur, de deux écrans d'affichage visuel et d'un bras haptique. Le dispositif haptique représente l'instrument rotatif du chirurgien-dentiste et, par retour de force sur la main, reproduit les sensations ressenties par le praticien au cours d'un soin dentaire. Ce simulateur propose un ensemble d'exercices avec différents niveaux de difficultés permettant aux étudiants d'être immergés dans leur future pratique et de se former aux gestes techniques.

Présentation de la méthode de recueil de données en phase préliminaire de la recherche. Cette étude préparatoire a concerné les étudiants de deuxième année de chirurgie dentaire de la Faculté d'Odontologie de Nantes. A ce stade de leur formation, ces étudiants n'ont jamais utilisé de simulateur conventionnel (« fantôme ») ou numérique. Ils ont été sélectionnés de façon aléatoire parmi les soixante-dix-sept étudiants de la promotion. Le premier protocole de cette recherche a permis aux étudiants de

tester quatre exercices proposés par le simulateur de réalité virtuelle. Ils ont d'abord utilisé *Virteasy Dental*® puis fantôme. Nous nous sommes intéressés aux ressentis et expériences vécues par les étudiants en utilisant ces simulateurs en se basant sur les modèles d'acceptabilité d'une technologie. Pour notre recueil de données, nous avons opté pour une approche quantitative par questionnaire auto administré. Nous avons collecté trente-huit réponses parmi les quarante étudiants inclus dans ce protocole. L'analyse des données a été réalisée selon un tri à plat. Les résultats sont présentés selon trois volets : la facilité d'utilisation du simulateur de réalité virtuelle, le ressenti des étudiants lors du passage du simulateur *Virteasy Dental*® au simulateur fantôme et les impressions générales quant à l'utilisation du simulateur de réalité virtuelle.

3 Résultats préliminaires

Facilité d'utilisation du simulateur Virteasy Dental ®. 50% des étudiants ont trouvé facile (47,4%) ou très facile (2,6%) l'utilisation du simulateur contre 39,5 % qui l'ont trouvé ni facile ni difficile. Seulement 10,5% des étudiants ont trouvé difficile (7,9%) ou très difficile (2,6%) l'utilisation de cet environnement. Les résultats montrent donc un avis partagé mais plutôt favorable quant à la facilité qu'ont les étudiants à utiliser le simulateur de réalité virtuelle. Concernant le ressenti de la préhension du bras haptique, 52,6% des étudiants déclarent que la manipulation de l'instrument n'est ni facile ni difficile. 15,8% des étudiants expriment une difficulté importante ou très importante et 31,6% trouvent cette préhension facile ou très facile. Ces résultats montrent un avis partagé quant à la manipulation du bras haptique. Notre enquête questionne également le ressenti des étudiants vis-à-vis de la position de travail imposée par le simulateur. Ce sont 57,9% des étudiants qui déclarent avoir mis du temps à adopter une position convenable contre 39,5% qui se sont adaptés facilement. Les questions relatives à la facilité d'exécution des exercices réalisés dans le cadre de cette étude montrent qu'il n'y a pas de lien entre les exercices réalisés et la facilité d'utilisation du simulateur de réalité virtuelle. D'après les premiers résultats, nous pouvons déterminer deux profils d'étudiants avec un premier ayant une facilité d'utilisation du simulateur et un second avec une difficulté plus ou moins importante. La manipulation de l'instrument haptique et l'adoption de la position de travail imposée par le simulateur semblent être deux facteurs ergonomiques diminuant la facilité d'utilisation de *Virteasy Dental*®.

Passage du simulateur Virteasy Dental® au simulateur conventionnel fantôme. 50% des étudiants déclarent avoir eu besoin d'un temps d'adaptation contre 42,1% qui n'ont pas eu de problème particulier. Cependant, les étudiants déclarent un niveau d'aisance plutôt satisfaisant et proche entre l'utilisation du bras haptique sur *Virteasy Dental*® (médiane=6) et la turbine sur fantôme (médiane=7) en faveur de l'outil conventionnel. De plus, 50% des étudiants déclarent que les exercices réalisés sur *Virteasy Dental*® les ont aidés à faire ceux demandés sur fantôme. Là encore, dans ce deuxième volet de résultats, deux profils d'étudiants se distinguent avec un premier n'ayant pas de difficulté à passer du simulateur de réalité virtuelle au simulateur conventionnel et un second qui éprouve des difficultés. L'ergonomie est différente entre les deux simulateurs que ce soit pour la position de travail ou l'instrument qui permet de réaliser les

soins dentaires. Cela peut donc impacter leur adaptation lors de la prise en main du simulateur.

Impressions générales à propos du simulateur Virteasy Dental®. 55,3% déclarent trouver ce simulateur utile dans l'apprentissage du geste de fraisage. 31,6% sont moyennement d'accord et 13,1% seulement ne sont pas d'accord. Lorsque nous avons demandé aux étudiants s'ils ont aimé utiliser le simulateur de réalité virtuelle, 81,6% ont répondu favorablement. 68,4% ont même répondu qu'ils souhaiteraient utiliser à nouveau cet environnement au cours de la formation. Nous pouvons tout de même souligner que 13,2 % des étudiants qui ont aimé utiliser le simulateur n'envisagent pas de l'utiliser à nouveau dans la formation. Peut-être, pouvons-nous mettre en cause « l'effet Waouh » [3] induit par l'introduction d'une nouvelle technologie sans pour autant être suffisant pour l'intégrer dans leur formation.

4 Conclusion et discussion

Nos résultats montrent que les étudiants ont en majorité aimé utiliser le simulateur *Virteasy Dental®*, l'ont trouvé utile dans l'apprentissage du geste de fraisage et souhaiteraient s'en servir à nouveau au cours de leur formation. Néanmoins deux profils se distinguent en ce qui concerne la facilité d'utilisation de l'environnement et l'ergonomie, ce qui pourrait avoir un impact sur son utilisation, son appropriation et donc sur son usage à long terme dans la formation des futurs chirurgiens-dentistes. Grâce à ces premières analyses, nous pouvons penser que le simulateur de réalité virtuelle peut effectivement trouver sa place au sein de la formation en Odontologie mais doit prendre en compte les difficultés rencontrées par les étudiants. La suite de notre programme de recherche consistera à approfondir les apports et les limites de cet environnement pour répondre aux mieux aux besoins des utilisateurs mais aussi du contexte de la formation.

Plus largement, cette recherche et ces résultats préliminaires questionnent l'intégration d'un artefact numérique dans un contexte d'enseignement-apprentissage. Ils reprennent l'idée exprimée par Jacquinot-Delaunay (2008) [4], selon laquelle l'insertion d'une technologie numérique, tel que le simulateur de réalité virtuelle, bouleverse le contexte de formation et relève d'un processus complexe qui entraine des modifications à tous les niveaux. Notre défi est de définir les éléments à prendre en compte lorsque l'on souhaite introduire un outil numérique en contexte d'enseignement-apprentissage.

Références

1. Basque, J. Doré, S.: Le concept d'environnement d'apprentissage informatisé. Revue de l'éducation à distance, vol 13, no 1, 40-56 (1998).
2. Charlier, B.: « Chapitre 2. Les environnements numériques d'apprentissage : quelques éléments d'intelligibilité pour la e-formation », Annie Jézégou éd., Traité de la e-formation des adultes. De Boeck Supérieur, 49-68 (2019).
3. Citton, Y.: L'initiation numérique : de l'entreprise des données à la grâce du commun. Être initié à l'heure des mutations anthropologiques, hal-02912313, (2017).
4. Jacquinot-Delaunay, G. Fichez, E.: L'université et les TIC : chronique d'une innovation annoncée. De Boeck, (2008).

Vers une analyse des pratiques informationnelles des acteurs impliqués dans le domaine des serious games

Mamoudou Ndiaye, Fabrice Pirolli, Raphaëlle Cretin-Pirolli

Le Mans Université, CREN, EA 2661 Le Mans, France
prenom.nom@univ-lemans.fr

Résumé. L'utilisation croissante du numérique dans l'éducation a entraîné des changements importants dans les méthodes et les pratiques d'enseignement. Les ressources numériques telles que les serious games constituent un élément important de cette évolution. Cet article traite de la nécessité d'analyser les pratiques informationnelles des acteurs impliqués dans la création et l'utilisation des jeux sérieux afin de trouver des solutions qui répondent à leurs besoins informationnels à travers une étude qualitative de leur environnement. Il questionne également l'intérêt d'améliorer les modèles de classification des jeux sérieux en Sciences de l'Information.

Mots-clés : Jeux sérieux, pratiques informationnelles, adaptation, jeux vidéo, modèles de classification des jeux vidéo

Abstract. The increasing use of digital technology in education has led to significant changes in teaching methods and practices. Digital resources such as serious games are an important part of this development. This article discusses the need to analyze the informational practices of the actors involved in the creation and use of serious games in order to find solutions that meet their informational needs through a qualitative study of their environment. It also questions the interest of improving the classification models of serious games.

Keywords: Serious games, information practices, adaptativity, video games, video games classification models

1 Introduction

L'avènement progressif du numérique dans le champ de l'éducation a engendré au cours des dernières décennies de profondes mutations dans les pratiques d'acteurs dont les méthodes et modalités d'enseignement ont été largement transformées. Déploiement de nouveaux dispositifs, de nouvelles approches, diversification et enrichissement des ressources pédagogiques ont largement contribué à ces mutations.

Sur le plan des ressources pédagogiques, le numérique a notamment, et de façon régulière, remis en lumière la question des apports potentiels des serious games aux dispositifs de formation. La conception, puis la mise en œuvre ou la gestion de ces

artefacts pédagogiques spécifiques font intervenir plusieurs catégories d'acteurs et mobilisent différentes disciplines (développeurs de jeux, enseignants, bibliothécaires et archivistes). Dans leurs contextes d'exercice respectifs, tous ces acteurs ont des besoins informationnels relatifs à l'identification, au repérage, à l'usage et à la communication de serious games. Dans le domaine de l'information-documentation plusieurs modèles de description ont été élaborés pour permettre à ces différents acteurs d'accéder à des ressources relatives aux jeux. Cependant, aucun de ces modèles ne fait l'unanimité quant à sa capacité effective à décrire le jeu sérieux dans sa spécificité.

C'est dans cette dynamique que notre projet de recherche doctorale se propose d'étudier ces dimensions sous l'angle de l'analyse des pratiques informationnelles des acteurs. Notre étude aidera à mieux comprendre comment les enseignants et les chercheurs naviguent dans le flot d'informations relatives aux ressources pédagogiques numériques et plus spécifiquement aux jeux sérieux.

2 Cadrage théorique

La délicate définition des serious games. Il n'existe pas de définition stabilisée et universellement admise des serious games. Clark Abt les conçoit comme « *des jeux explicitement et intentionnellement conçus à des fins éducatives, et qui ne sont pas principalement destinés au divertissement [1].* »

Cette définition rejoint celle de Michael et Chen, concepteurs de jeux sérieux qui les définissent comme « *tout jeu dont la finalité première est autre que le divertissement* ». L'expression désigne donc un jeu sur tout type de support avec comme particularité d'associer la dimension amusement avec la fonction utilitaire [2].

Nos travaux s'inscrivent dans la perspective énoncée par Julien Alvarez, spécialisé dans les Technologies de l'Information et de la Communication pour l'Éducation (TICE), qui les définit de façon plus précise comme « *Application informatique, dont l'objectif est de combiner à la fois des aspects sérieux (Serious) tels, de manière non exhaustive, l'enseignement, l'apprentissage, la communication, ou encore l'information, avec des ressorts ludiques issus du jeu vidéo (Game)[3]*.» C'est sur la base de cette définition que nous questionnons les pratiques effectives des différents acteurs dans le cadre de notre projet de recherche.

Une approche des pratiques informationnelles sous l'angle de l'information-documentation. De nombreux facteurs peuvent influencer les pratiques d'une personne en situation de recherche d'information, notamment ses connaissances et sa compréhension d'un sujet, ses objectifs personnels et professionnels, les ressources et les outils à sa disposition, ainsi que les normes sociales et culturelles qui façonnent ses comportements et ses attentes. Pour mieux appréhender ces différents facteurs, il est nécessaire de faire une analyse de ces pratiques, dites pratiques informationnelles, au sens défini par Chaudiron et Ihadjadene : « *la manière dont un ensemble de dispositifs, de sources formelles ou non, d'outils, de compétences cognitives sont effectivement mobilisés, par un individu ou un groupe d'individus, dans les différentes situations de production, de recherche, d'organisation, de traitement, d'usage, de partage et de communication de l'information [4]*.» En plus d'aider à comprendre les pratiques informationnelles d'acteurs, notre analyse pourra contribuer à la conception de systèmes, services et d'outils d'information qui soutiennent des activités d'éducatives.

3 De l'intérêt de l'analyse des pratiques d'acteurs pour l'adaptation des serious games

Les serious games tels qu'on les appréhende aujourd'hui, c'est-à-dire en tant que concept caractérisant un produit numérique [5], ont été travaillés au début des années 2000. C'est dans ce contexte qu'on assiste à l'émergence des jeux sérieux en tant que branche des jeux vidéo. Des jeux sont alors conçus dans un but autre que le seul divertissement [2]. Le concept commence alors à susciter l'intérêt dans le milieu académique [5], de nombreux enseignants se lancent à la recherche de serious games utilisables en classe. De nombreux chercheurs et enseignants participent à la création de serious games pour les élèves, entraînant une prolifération de ces jeux. Pour produire ces ressources pédagogiques, les enseignants s'intéressent à l'adaptation des serious games à leurs cours en proposant, adaptant et sélectionnant des informations en fonction des objectifs et du profil de l'apprenant. Connaître leurs besoins et pratiques informationnelles est donc essentiel pour mieux les satisfaire.

4 Les enjeux de la description des serious games

Dans le champ des Sciences de l'Information et de la documentation, plusieurs modèles ont été élaborés pour décrire les ressources du jeu vidéo, en tant qu'unité documentaire, afin d'en faciliter l'accès, l'usage et la réutilisation dans un contexte personnel ou professionnel. L'intérêt des services d'information (bibliothèques, archives, centres de documentation) pour les ressources du jeu vidéo a fortement contribué au développement de ces modèles de description. Cependant, il n'existe toujours pas de normes communément acceptées pour décrire les jeux vidéo. Celles qui préexistaient dans les fonds documentaires ne permettent pas de décrire efficacement le jeu car elles demeurent plus adaptées aux ressources documentaires et ne prennent pas en compte les spécificités des jeux vidéo [6]. Par exemple, à la *Library of Congress* (États Unis), les jeux vidéo sont catalogués à l'aide de métadonnées conçues pour les films et les enregistrements sonores. Cette solution soulève des questions sur les défis que les chercheurs pourraient rencontrer en essayant de trouver des informations relatives aux jeux sur des bases de données ainsi organisées [7]. Elles sont insuffisantes pour les jeux vidéo et ne les décrivent pas de manière satisfaisante pour les chercheurs [8]. Le modèle FRBR (*Functional Requirements for Bibliographic Records*) donne une vision synthétique et simplifiée des catalogues de Bibliothèque. Souvent utilisé pour la description des jeux vidéo, il présente des limites en raison de la complexité de ces derniers et de l'absence de certaines de ses caractéristiques [6]. Les entités *Oeuvre, Expression, Manifestation et item* ne sont pas faciles à distinguer pour les jeux vidéo [9]. Jugeant le modèle FRBR inapproprié, Jacob Jett propose le "modèle conceptuel pour les Jeux Vidéo" (*Conceptual model for Video Games and Interactive Media*) [6]. Ce Modèle tente de refléter la manière dont les utilisateurs, tels que les joueurs, les collectionneurs et les chercheurs comprennent les jeux vidéo et les relations entre eux. Inspiré du Modèle de Jacob Jett, le *Mediated Vocabulary For Video Game Research* [10] explique comment la combinaison de différentes sources d'information, telles que les métadonnées décrivant les jeux et d'autres détails connexes, peuvent être utilisées pour reconstruire l'historique de sortie des jeux vidéo.

Le modèle qui servira de base à notre étude sera celui de Julien Alvarez appelé G/P/S [11]. Celui-ci est basé sur trois critères *Gameplay*, *Permet de*, *Secteur*, permettant de prendre en compte les dimensions ludique et sérieuse des jeux. Il a été élaboré à partir des systèmes de classification mentionnés ci-dessus.

L'étude des pratiques informationnelles des acteurs que nous entendons mener mobilisera ces différents concepts, en vue de confronter modèles théoriques et réalités des usages et accommodations dans le champ éducatif.

5 Perspectives de recherche

Notre thèse vise à étudier les pratiques des acteurs des serious games éducatifs en menant une étude qualitative auprès de concepteurs et usagers pour caractériser leur environnement informationnel. Les objectifs incluent l'analyse des pratiques informationnelles, l'enrichissement des modèles de classification et l'évaluation de l'impact pédagogique des serious games. Nous réaliserons des entretiens avec concepteurs, usagers et gestionnaires de corpus de serious games. La classification des jeux sérieux soulève également la problématique de leur patrimonialisation, déjà abordée dans les années 1990 . En France, la Bibliothèque Nationale de France (BnF) collecte les logiciels édités, distribués ou importés, dont les jeux vidéo, suite à la loi de 1992 sur le dépôt légal. La BnF, disposant d'un fonds de jeux vidéo incluant des jeux sérieux, servira de cadre institutionnel à notre étude, complétant les travaux de terrain dans le domaine de l'éducation.

Références

1. Abt, C.C.: Serious Games. University Press of America (1987).
2. Michael, D., Chen, S.: Serious Games: Games That Educate, Train, and Inform. (2006).
3. Alvarez, J.: Du jeu vidéo au serious game, https://hal.archives-ouvertes.fr/tel-01240683, (2007).
4. Chaudiron, S., Ihadjadene, M.: De la recherche de l'information aux pratiques informationnelles. Études de communication. langages, information, médiations. (2010).
5. Lavigne, M.: Les faiblesses ludiques et pédagogiques des serious games.
6. Jett, J., Sacchi, S., Lee, J.H., Clarke, R.I.: A conceptual model for video games and interactive media. Journal of the Association for Information Science and Technology. 67, 505–517 (2016).
7. Lee, J.H.: Empirical Evaluation of Metadata for Video Games and Interactive Media. (2015).
8. Owens, T.: Yes, The Library of Congress Has Video Games: An Interview with David Gibson | The Signal, //blogs.loc.gov/thesignal/2012/09/yes-the-library-of-congress-has-video-games-an-interview-with-david-gibson/, last accessed 2023/01/12.
9. McDonald, C., Schmalz, M., Monheim, A., Keating, S., Lewin, K., Cifaldi, F., Lee, J.H.: Describing, organizing, and maintaining video game development artifacts. Journal of the Association for Information Science and Technology. 72, 540–553 (2021).
10. Hoffmann, T.: A Mediated Vocabulary for Video Game Research. 11.
11. Djaouti, D., Alvarez, J., Jessel, J.-P., Djaouti, D., Alvarez, J., Jessel, J.-P.: Classifying Serious Games: The G/P/S Model, https://www.igi-global.com/gateway/chapter/www.igi-global.com/gateway/chapter/52492, last accessed 2023/01/11.=

Vers une analyse temporelle des processus rédactionnels des élèves

Léo Nebel[1][0000−0001−5859−265X], François Bouchet[2][0000−0001−9436−1250] et Denis Alamargot[3][0000−0001−9807−3723]

[1] EvidenceB leo-n@evidenceb.com
[2] Sorbonne Université, CNRS, LIP6, F-75005 Paris, France
francois.bouchet@lip6.fr
[3] Université de Poitiers - INSPE de l'académie de Poitiers Laboratoire CHArt-UPEC
denis.alamargot@univ-poitiers.fr

Résumé. L'évaluation formative des productions écrites dans un cadre scolaire est un sujet complexe. Cet exercice mobilise chez les élèves de nombreuses compétences, dont notamment la maîtrise des processus rédactionnels. De nombreux outils existent déjà pour aider les professeurs dans leurs évaluations, voire dans certains cas pour automatiser l'évaluation d'un texte. Cependant, il n'existe pas d'outil à notre connaissance qui exploite des traces temporelles afin d'analyser la maîtrise des processus rédactionnels par les élèves, en parallèle des méthodes d'évaluation statiques plus classiques. Nous examinons ici les apports potentiels d'une telle analyse temporelle dans le cadre d'une évaluation formative d'écrits d'élèves.

Mots-clé: traitement automatique de la langue naturelle · processus rédactionnels · analyse temporelle · révision · évaluation formative

Abstract. The formative evaluation of written work in a school setting is a complex subject. This exercise mobilizes many skills for students, including the mastery of writing processes. Many tools already exist to help teachers in their evaluations, even in some cases to automate the evaluation of a text. However, to our knowledge, there are no tools that exploit time traces to analyze students' mastery of writing processes, in parallel with more traditional static assessment methods. We examine here the potential contributions of such a temporal analysis in the context of a formative assessment of students' writing.

Keywords: natural language processing · writing processes · temporal analysis · revising process · formative assessment

1 Introduction

L'évaluation représente une grande partie du temps du professeur. À travers le développement des usages d'outils numériques au sein de milieux éducatifs,

son automatisation devient de plus en plus courante. Cette automatisation est toutefois plus compliquée dans le cadre d'exercices incluant du texte libre. Des modèles basés sur des données statistiques existent depuis longtemps, d'autres plus récents se fondent aussi sur des méthodes de *deep learning* [8]. Les corrections proposées par ces modèles peuvent porter sur le style du texte, sa forme, son sens et/ou sa pertinence. Ces corrections correspondent à celles réalisées par le professeur qui n'a accès qu'à la production finale d'un élève. Mais au-delà du gain de temps apporté par les outils numériques, ils permettent de suivre le déroulé de l'activité de rédaction de l'élève rendant possible une analyse automatique. On peut alors se poser les questions de recherche suivantes : (1) cette analyse automatique dynamique permet-elle d'enrichir la qualité de l'évaluation ? (2) permet-elle d'améliorer aussi les rétroactions fournies par l'EIAH (Environnement Informatique pour l'Apprentissage Humain), notamment les rétroactions formatives et motivationnelles ?

2 État de l'art de l'évaluation de texte d'élèves

De nombreux travaux ont déjà été réalisés dans le champ de l'évaluation automatique d'essais (AES pour *Automated Essay Scoring*, au caractère sommatif) et d'écrits (AWE pour *Automated Writing Evaluation*, au caractère formatif) [7].

Les premiers systèmes d'évaluation se concentraient sur des critères liés au style de la production écrite. Ils s'intéressaient notamment à la diversité du vocabulaire à travers des indicateurs comme le ratio *Type-Token*, la correction orthographique ou la complexité syntaxique [3]. Ces évaluations de forme sont rendues possibles depuis un certain temps grâce à différentes techniques de traitement automatique du langage naturel (tokenisation, analyse morpho-syntaxique, dictionnaires...). Des systèmes plus récents évaluent aussi le fond et la sémantique des productions. La vectorisation des mots, des phrases et des textes à l'aide de modèles pré-entraînés basés sur l'architecture Transformers tels que BERT [5] ou d'autres modèles similaires [4] ont permis ces nouveaux résultats.

Peu de systèmes mêlent ces deux modes d'évaluation (fond et forme). Si les modèles de *deep learning* prennent indirectement en compte la forme dans leur notation (des corrélations entre forme et fond existant dans un nombre important de productions), cette évaluation statique ne permet pas de comprendre la stratégie rédactionnelle de l'élève et l'enchaînement des processus dans le temps.

Il est ainsi possible d'ajouter une troisième dimension à cette évaluation : l'analyse temporelle. La plupart des systèmes d'évaluation se limitent à évaluer l'état final du texte (le produit). Or dans un objectif d'évaluation formative, l'évaluation de la maîtrise du décours et de la gestion temporelle est essentielle [2]. De nombreux travaux ont montré que les bons écrivains mobilisaient et enchaînaient des processus de rédaction spécifiques [1]. À travers les outils pédagogiques numériques, nous pouvons mettre en place des systèmes retraçant de manière précise l'évolution temporelle d'un texte. À notre connaissance, peu de modèles ajoutent cette composante dans leur évaluation. Même s'il existe des recherches où des feedbacks ont été donnés à partir d'analyses temporelles [6], ces

dernières ne proposent pas de modèles automatisés permettant cette évaluation ni une évaluation croisée avec les critères plus classiques discutés ci-dessus.

3 Contexte et modélisation

Nous proposons ainsi de tester une évaluation dynamique afin d'observer les avantages que cette dernière pourrait apporter dans la notation et le feedback écrit proposé aux élèves. Cette évaluation aura un double caractère dynamique : (1) à travers l'analyse de l'évolution du texte mais aussi (2) à travers le suivi continu de l'élève et la mesure de ses progrès, cette deuxième partie n'étant pas traitée ici. Nous nous appuyons sur un modèle d'exercice dans lequel on donne à l'élève une introduction et une conclusion, et où on lui demande de faire le lien entre les deux en rédigeant un texte. Cet exercice est réalisé par des élèves de lycée dans un exerciseur plus général visant la remise à niveau sur différentes compétences. L'outil développé évalue le texte rédigé selon 6 critères : (1) la sémantique, (2) la syntaxe, (3) l'orthographe, (4) le vocabulaire, (5) la fluence, (6) la révision. Les quatre premiers critères s'évaluent traditionnellement uniquement à partir de l'état final (statique) du texte tandis que les deux derniers nécessitent de suivre son évolution (dynamique). Cette évaluation de la dynamique nécessite une collecte des frappes claviers, des temps associés et de la position du curseur.

Concernant la fluence, l'enjeu est d'évaluer la vitesse d'écriture de l'élève comparée à sa vitesse d'écriture "habituelle". Cela peut se faire à partir des données collectées et d'un comptage dynamique des mots écrits via une tokenisation. Pour le critère de révision, l'enjeu consiste à construire un score cohérent à partir de multiples indicateurs pour quantifier la tendance à la révision de l'élève. Parmi ces indicateurs on trouve le nombre et le temps médian de pause, le ratio de mots produits par rapport aux mots processés (c'est-à-dire les mots qui ont été effacés par la suite), ou encore la distance sémantique entre les différents états du texte après chaque phase de relecture. Le score de révision se calcule par une moyenne pondérée de l'ensemble des indicateurs. Cette pondération se fait de manière équilibrée mais sera sujette à un calibrage à l'avenir. À partir de ce score, un retour global sera donné à l'élève, ainsi que des retours plus précis sur chacun des indicateurs afin de ne pas perdre en information.

Au début de notre modélisation, nous avons collecté des données auprès d'un public varié, constitué d'une cinquantaine de réponses, accompagnées des traces clavier permettant d'effectuer nos analyses dynamiques. Ces données nous ont permis d'établir une première version d'un système expert qui a pour prochain objectif d'être ajusté auprès d'un public cible d'élèves plus important.

4 Prochaines étapes

L'étape immédiate à venir sera de valider la première version de notre modèle. Une expérimentation va être réalisée dans ce cadre auprès d'un public cible d'élèves. La progression des élèves à travers plusieurs sessions sera mesurée et

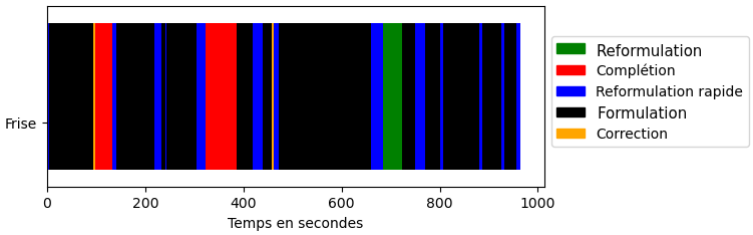


Fig. 1. Représentation temporelle des transitions d'états du texte d'un élève

comparée selon les différents cas envisagés. On pourra ainsi évaluer l'influence de l'analyse dynamique en complément de l'analyse statique par rapport à une simple analyse statique.

À moyen terme, il serait intéressant de construire à partir de nos différents indicateurs, des représentations temporelles de l'évolution du texte (cf. Figure 1). Ce découpage permettrait de mieux rendre compte des stratégies rédactionnelles existantes et de leur influence sur la qualité générale du texte. Dans un second temps, une collecte massive de traces pourra être effectuée à travers une application éducative. Ces données, une fois étiquetées, nous permettront de construire automatiquement des représentations temporelles des réponses d'élèves potentiellement plus précises que celles construites à partir de nos simples indicateurs. Des expérimentations permettront de confirmer cette hypothèse.

References

1. Alamargot, D., Lebrave, J.L.: The Study of Professional Writing. European Psychologist **15**(1), 12–22 (Jan 2010). https://doi.org/10.1027/1016-9040/a000001
2. Alamargot, D., Plane, S., Lambert, E., Chesnet, D.: Using eye and pen movements to trace the development of writing expertise: case studies of a 7th, 9th and 12th grader, graduate student, and professional writer. Reading and Writing **23**(7), 853–888 (Aug 2010). https://doi.org/10.1007/s11145-009-9191-9
3. Blache, P.: Un modèle de caractérisation de la complexité syntaxique. traitement automatique des langues naturelles pp. 1–10 (2010)
4. Chang Lu, M.C.: Integrating deep learning into an automated feedback generation system for automated essay scoring. Proceedings of The 14th International Conference on Educational Data Mining (EDM 2021) pp. 573–579 (2021)
5. Devlin Jacob, Chang Ming-Wei, L.K., Kristina, T.: Bert: Pre-training of deep bidirectional transformers for language understanding. pp. 4171–4186. Association for Computational Linguistics, Minneapolis, Minnesota (2019)
6. Nina Vandermeulen, Elke Van Steendam, S.D.M., Rijlaarsdam, G.: Writing process feedback based on keystroke logging and comparison with exemplars: Effects on the quality and process of synthesis texts. Written communication **40**, 90–144 (2023)
7. Ramesh Dadi, S.K.S.: An automated essay scoring systems: a systematic literature review. Artificial Intelligence Review **55**, 2495–2527 (2022)
8. Uto, M.: A review of deep-neural automated essay scoring models. Behaviormetrika **48**, 459–484 (2021)

Analyse du comportement de l'apprenant dans l'apprentissage en ligne et adaptatif

Omar Oustous[1], Adel Ihichr[1], Younès El Bouzekri El Idrissi[1], Ayoub Ait Lahcen[1]

[1] Laboratoire des Sciences de l'Ingénieur, ENSA, Université Ibn Tofail, Kénitra, Maroc
prenom.nom@uit.ac.ma

Résumé. Ces dernières années, le développement et la complexité croissante des systèmes d'apprentissage en ligne ont créé un intérêt de modéliser les caractéristiques du comportement des apprenants. Ainsi, le présent article vise à fournir une revue systématique de la littérature sur l'analyse du comportement de l'apprenant dans les environnements d'apprentissage en ligne et adaptatif.

Mots-clés : comportement de l'apprenant, apprentissage en ligne, apprentissage adaptatif, e-learning, EIAH

Abstract. In recent years, the development and complexity increase of e-learning systems have created an interest in modeling the characteristics of students' behavior in e-learning environment. In this context, the present paper aims to provide a systematic review of the literature about the analysis of student behavior in online and adaptive learning environments.

Keywords: learner behavior, online learning, adaptive learning, e-learning, EIAH

1 Introduction

L'utilisation croissante des systèmes hypermédias, accessibles par une grande variété d'utilisateurs, a créé un besoin de caractériser dynamiquement le comportement de l'utilisateur. Cette caractérisation en ligne des apprenants constitue une étape préliminaire au développement de systèmes d'aide adaptatifs. En effet, l'adaptabilité tient compte des différences individuelles par le biais de la modélisation de l'apprenant (modèle de l'apprenant), qui a été étendue pour prendre en considération un éventail plus large de caractéristiques du comportement de l'apprenant.

2 Analyse des caractéristiques du comportement de l'apprenant

Ces dernières décennies, les systèmes d'apprentissage adaptatif ont reçu une attention importante et un intérêt particulier. Certains chercheurs ont examiné les conceptions, le

contexte, les stratégies et les technologies des systèmes d'apprentissage adaptatif [1]. Les résultats de cette étude ont montré qu'il y a eu six évaluations systématiques réalisés sur l'apprentissage adaptatif entre 2009 et 2018. Ces analyses se sont concentrées sur les critères d'adaptation, les caractéristiques distinctes d'un système d'apprentissage adaptatif telles que les différences individuelles, les traits personnels et l'efficacité du système. Les résultats ont montré que la majorité des études (67,2 %) ciblaient le cognitif comme variable de résultat et utilisaient un design expérimental. Bien que plus modestes, il convient de noter que plus d'un tiers des études ciblaient les résultats affectifs (37,7 %) et encore plus les résultats comportementaux (41 %).

Les principales caractéristiques du comportement de l'apprenant couvertes par ce présent article sont les suivantes (par ordre alphabétique) : Affect/Emotions, Autorégulation, Collaboration, Engagement et Motivation. Ainsi, nous avons ajouté et dédié une sous-section « Gamification » pour montrer les avancées dans le domaine des jeux éducatifs en termes de modélisation du comportement de l'apprenant.

Affects/Emotions. Récemment, des variables affectives spécifiques au domaine de l'éducation ont été évaluées dans les environnements d'apprentissage en ligne, telles que l'ennui, la frustration, la joie, la honte et l'incertitude. En effet, une nouvelle méthode basée sur la détection des visages à marge maximale pour l'analyse des états d'affect positifs et négatifs des apprenants en utilisant leurs expressions faciales a été proposée [2]. En outre, [3] ont proposé une conception d'un modèle d'apprenant pour adapter le feedback formatif en fonction des états affectifs des apprenants. Les résultats suggèrent que le soutien basé sur l'affect contribue à réduire l'ennui et le comportement hors tâche, et peut avoir un effet positif sur l'apprentissage.

Autorégulation. Être capable de réguler son propre apprentissage est considéré par les psychologues de l'éducation comme la clé d'un apprentissage réussi. L'apprentissage autorégulé fait référence à des situations d'apprentissage dans lesquelles les apprenants fixent leurs propres objectifs d'apprentissage [4]. Un bon retour d'information peut renforcer la capacité des élèves à autoréguler leur propre performance et constitue donc un aspect important à prendre en compte dans l'étude de l'apprentissage autorégulé. L'évaluation des progrès de l'apprentissage sont essentiels pour un apprentissage autorégulé réussie.

Collaboration. Des chercheurs ont analysé des modèles d'interaction collaborative d'équipes d'apprentissage composées d'apprenants [5]. L'étude de recherche a principalement montré que les catégories de communication avec le plus grand nombre d'activités étaient l'organisation du travail, l'initiation d'activités, le feedback donné et l'interaction sociale. [6] ont démontré que l'apprentissage en petit groupe avec l'apprentissage collaboratif assisté par ordinateur a des effets plus positifs sur les processus cognitifs et les résultats affectifs d'un apprenant que l'apprentissage individuel. Des chercheurs ont établi une méta-analyse sur les effets de l'apprentissage collaboratif assisté par ordinateur [7]. Ils ont trouvé que la collaboration a eu des effets positifs significatifs sur le gain de connaissances, l'acquisition de compétences et les perceptions des apprenants.

Engagement. L'engagement comporte trois composantes : comportemental, émotionnel et cognitif. Il est associé à des résultats scolaires positifs, y compris la réussite

et la persévérance à l'école, il est plus élevé dans les classes où les enseignants et les pairs apportent leur soutien, où les tâches sont stimulantes et authentiques et où les choix sont possibles [8]. Les chercheurs ont souligné une absence d'un consensus autour de la définition du concept et un manque de distinction entre l'état de l'engagement, les facteurs qui l'influencent, et ses conséquences immédiates à court terme et à plus long terme[9]. L'auto-efficacité académique, la présence de l'enseignement et l'utilité perçue ont des effets directs significatifs sur l'engagement d'apprentissage [10].

Gamification. Compte tenu de l'évolution des besoins d'apprentissage de l'apprenant, l'apprentissage avec des jeux numériques au lieu de la méthode traditionnelle qui est devenue ennuyeuse et non motivante est devenu fortement recommandé [11]. [12] ont développé un réseau bayésien dynamique basé sur les expressions corporelles des élèves dans un jeu éducatif, Prime Climb. Dans ce jeu, un agent pédagogique fournissait un soutien en fonction de l'état émotionnel des élèves détecté par le système et de leur objectif personnel. Les deux chercheurs [13] ont passé en revue la littérature relative à la pédagogie des jeux et de la simulation. Ils ont indiqué que les jeux et/ou les simulations ont un impact positif sur les objectifs d'apprentissage. Ainsi, les chercheurs ont trouvé que les jeux éducatifs sont de plus en plus acceptés comme un outil d'apprentissage utile pour générer des environnements éducatifs plus engageants [14].

Motivation. « La motivation peut être comprise comme le " pourquoi " ou la raison derrière une réponse physique, émotionnelle ou cognitive donnée » [15]. Les chercheurs ont développé un modèle permettant de détecter plusieurs aspects de la motivation, notamment le désir de contrôle, de défi et d'indépendance. Les stratégies pédagogiques qui visaient à favoriser la motivation des apprenants étaient souvent mises en œuvre dans le mode en ligne [16]. Les exemples comprenaient la mise en œuvre d'activités interactives en ligne, telles que des quiz, des jeux, des puzzles et des exercices flash.

3 Conclusion et perspectives

Les chercheurs ont analysé les différentes caractéristiques du comportement de l'apprenant selon des modélisations, des techniques de mesure et des démarches différentes. Ils ont soulevé l'utilisation fréquente des questionnaires comme moyen de mesure pour des échantillons de petites tailles, ce qui pose des problèmes de validité, de fiabilité et même de généralisation des résultats.

La prise en compte des caractéristiques du comportement dans la mise en place des interventions pédagogiques a montré des effets positifs sur l'expérience de l'apprenant avec le système et sur les résultats d'apprentissage. Ainsi, pour une meilleure exploitation des différences comportementales des apprenants, des études expérimentales et multidimensionnelles sur des grandes échelles seront sollicitées pour démontrer les avantages et les effets positifs de la modélisation du comportement de l'apprenant dans la mise en œuvre des systèmes d'apprentissage en ligne et adaptatif.

Remerciements : Ce travail a été soutenu par le MESRSI, l'ADD et le CNRST du Maroc dans le cadre du programme Al-Khawarizmi, projet n°451/2020 (Smart Learning).

Références

1. Martin, F., Chen, Y., Moore, R.L., Westine, C.D.: Systematic review of adaptive learning research designs, context, strategies, and technologies from 2009 to 2018. Education Tech Research Dev. 68, 1903–1929 (2020).
2. Gupta, S.K., Ashwin, T.S., Guddeti, R.M.R.: Students' affective content analysis in smart classroom environment using deep learning techniques. Multimed Tools Appl. 78, 25321–25348 (2019).
3. Grawemeyer, B., Mavrikis, M., Holmes, W., Gutiérrez-Santos, S., Wiedmann, M., Rummel, N.: Affective learning: improving engagement and enhancing learning with affect-aware feedback. User Model User-Adap Inter. 27, 119–158 (2017).
4. van Seters, J.R., Ossevoort, M.A., Tramper, J., Goedhart, M.J.: The influence of student characteristics on the use of adaptive e-learning material. Computers & Education. 58, 942–952 (2012).
5. Serçe, F.C., Swigger, K., Alpaslan, F.N., Brazile, R., Dafoulas, G., Lopez, V.: Online collaboration: Collaborative behavior patterns and factors affecting globally distributed team performance. Computers in Human Behavior. 27, 490–503 (2011).
6. Lou, Y., Abrami, P.C., d'Apollonia, S.: Small Group and Individual Learning with Technology: A Meta-Analysis. Review of Educational Research. 71, 449–521 (2001).
7. Chen, J., Wang, M., Kirschner, P.A., Tsai, C.-C.: The Role of Collaboration, Computer Use, Learning Environments, and Supporting Strategies in CSCL: A Meta-Analysis. Review of Educational Research. 88, 799–843 (2018).
8. Fredricks, J.A., Blumenfeld, P.C., Paris, A.H.: School Engagement: Potential of the Concept, State of the Evidence. Review of Educational Research. 74, 59–109 (2004).
9. Kahu, E.R.: Framing student engagement in higher education. Studies in Higher Education. 38, 758–773 (2013).
10. Jung, Y., Lee, J.: Learning Engagement and Persistence in Massive Open Online Courses (MOOCS). Computers & Education. 122, 9–22 (2018).
11. Denden, M., Tlili, A., Essalmi, F., Jemni, M.: Implicit modeling of learners' personalities in a game-based learning environment using their gaming behaviors. Smart Learn. Environ. 5, 29 (2018).
12. Conati, C., Maclaren, H.: Empirically building and evaluating a probabilistic model of user affect. User Model User-Adap Inter. 19, 267–303 (2009).
13. Vlachopoulos, D., Makri, A.: The effect of games and simulations on higher education: a systematic literature review. Int J Educ Technol High Educ. 14, 22 (2017).
14. Saleem, A.N., Noori, N.M., Ozdamli, F.: Gamification Applications in E-learning: A Literature Review. Tech Know Learn. 27, 139–159 (2022).
15. Hew, K.F.: Promoting engagement in online courses: What strategies can we learn from three highly rated MOOCS: Engagement: lessons from MOOCs. Br J Educ Technol. 47, 320–341 (2016).
16. Boelens, R., De Wever, B., Voet, M.: Four key challenges to the design of blended learning: A systematic literature review. Educational Research Review. 22, 1–18 (2017).

Système de recommandation de cours basé sur les soft skills : Une approche utilisant les algorithmes génétiques

Luis Alberto Pinos Ullauri[1,2,3][0000−0002−3694−3487], Alexis Lebis[1], Abir Karami[4], Mathieu Vermeulen[1] Anthony Fleury[1], and Wim Van Den Noortgate[2,3]

[1] IMT Nord Europe, Centre for Digital Systems, Lille, France
[2] KU Leuven, Faculty of Psychological and Educational Sciences, Leuven, Belgium
[3] imec ITEC KU Leuven, Leuven, Belgium
[4] Université Catholique de Lille, Lille, France

Résumé L'importance croissante prise ces dernières années par les soft skills dans l'enseignement supérieur a encouragé les recherches sur leur relation avec les activités pédagogiques. Ce travail s'intéresse à l'incidence des soft skills sur la recommandation d'un ensemble de cours aux étudiants en fonction à la fois du profil de formation souhaité et des attentes du monde professionnel. Trois fonctions d'évaluation et deux méthodes de compensation sont proposées pour simuler différentes réalités socio-professionnelle et exprimer la différence entre la maîtrise des soft skills estimés heuristiquement et le score attendu à la fin du programme.

Keywords: Systèmes de recommandation· Soft Skills· Algorithmes Génétiques · Modèles Mixtes Linéaires · Fitness

Abstract. The increasing attention regarding soft skills in higher education has promoted research to study its relation with pedagogical activities. This work presents a framework to recommend a set of courses to students based on their soft skills considering their expected profile at the end of an academic program. Three score functions and two compensation methods are proposed to handle the difference between the heuristically estimated proficiency and the minimum expected proficiency of their soft skills by the end of the program.

Keywords: Recommender Systems· Soft Skills· Genetic Algorithms · linear mixed models · fitness

1 Introduction

L'intérêt pour les soft skills a été au départ porté principalement par l'industrie mais a maintenant une place importante dans les établissements d'enseignement supérieur (EES). La conceptualisation et la formalisation du terme «soft skills» font l'objet de débats entre les chercheurs. Cependant, une interprétation générale, telle que présentée par [1], les décrit comme un groupe de compétences non techniques, intrapersonnelles et interpersonnelles que nous pouvons utiliser

en diverses situations. De leur côté, [3] ont réalisé trois études différentes dans lesquelles ils ont constaté que la demande de soft skills variait en fonction des disciplines des étudiants, soutenant ainsi l'idée des profils soft skills différents, comme le soutient [2]. En plus, la notion de soft skills peut s'entremêler avec celle de compétences, néanmoins, les compétences peuvent également faire référence à des aptitudes techniques, alors que les soft skills sont flexibles et multidisciplinaires (*i.e.* gestion du stress, leadership). Récemment, [5] a proposé un système de recommandation (SR) pour individualiser le cursus des étudiants par rapport aux compétences et le cadre formel associé. Néanmoins, cette recherche fondamentale ne traîte pas du caractère particulier des soft skills, ni de l'intérêt que peuvent avoir les étudiants concernant les cours proposés et l'effet associé. Notre problématique est alors de pouvoir permettre la recommandation de parcours ancrés dans ces singularités.

2 Contributions et Propositions

Un scénario hypothétique, basé sur une situation réelle dans un ESS en France, est présenté afin d'illustrer le problème. Les travaux de [5] ont prouvé que le problème de recommandation de parcours individualisé – où un étudiant dans un cursus doit suivre un nombre fini de cours dans des périodes temporels spécifiques comme des semestres – est au moins NP-Complet. Les auteurs proposent une modélisation générique : les cours sont piochés dans un catalogue de cours en autorisant les combinaisons avec répétition. Pour limiter la complexité du problème, nous avons décidé de nous placer dans un cas de SR où les combinaisons sont sans répétition. Dans notre scénario, les étudiants doivent suivre $n = 11$ cours parmi $m = 16$ (C_{16}^{11}.) options possibles et quatre spécialisations différentes sont envisagées. Nous supposons également que les cours ne sont pas spécifiques à une spécialisation et nous considérons pour le moment trois soft skills indépendants. Les soft skills sont traités comme des valeurs continues et pouvaient être mesurés avec IRT [4]. Chaque spécialisation définie des attentes (*i.e.* seuils) décrivant la maîtrise des soft skills attendue, montré dans le tableau 1.

Table 1 – Définition des seuils par spécialisation

Spécialisation	Soft Skill 1	Soft Skill 2	Soft Skill 3
Spéc. 1	0	0	0
Spéc. 2	0.5	0	-1
Spéc. 3	0	0.5	-1
Spéc. 4	-1	0	0.5

2.1 Génération de données

Nous proposons le modèle présenté dans l'équation 1 pour générer la maîtrise du soft skill i par l'étudiant s au semestre t. β_{i0} représente la maîtrise initiale moyenne sur le soft skill i d'un étudiant moyen. β_{ic} représente l'effet fixe du cours c sur le soft skill i. α_{ics} représente un effet aléatoire de l'étudiant s vers le cours c et le soft skill i. En outre, I_{cst} se rapporte à l'intérêt de l'étudiant s pour le choix du cours c au semestre t. De plus, θ_{is} représente l'effet aléatoire

de l'étudiant s dans le développement du soft skill i. Et, enfin, ϵ_i est une erreur résiduelle du soft skill i.

$$SSkill_{ist} = \beta_{i0} + \sum_{c=1}^{16}((\beta_{ic} + \alpha_{ics})D_{cst} + I_{cst}) + \theta_{is} + \epsilon_i \quad (1)$$

où D_{cst} est une variable fictive qui est égale à 1 si l'étudiant s a suivi le cours c au semestre t. Nous avons utilisé ce modèle pour produire des jeux de données simulés sur 1000 étudiants, du début jusqu'à bout d'un programme de 4 semestres. Les distributions de paramètres et les effets sont indiqués dans le tableau 2.

TABLE 2 – Distributions et valeurs des paramètres

Paramètre	Distribution	Effet
β_{i0}	-	(1.5,1,0.5)
α_{ics}	$N(\mu = 0, \sigma_{i=1} = 0.05, \sigma_{i=2} = 0.75, \sigma_{i=3} = 0.1)$	-
I_{cst}	$N(\mu = 0, \sigma = 0.1)$	-
θ_{is}	$N(\mu = 0, \sigma_{i=1} = 0.15, \sigma_{i=2} = 0.2, \sigma_{i=3} = 0.25)$	-
ϵ_i	$N(\mu = 0, \sigma_{i=1} = 0.2, \sigma_{i=2} = 0.15, \sigma_{i=3} = 0.1)$	-

2.2 Algorithmes génétiques et Fitness

Pour trouver des solutions à ce problème de recommandation basées sur les soft skills et l'intérêt de l'étudiant, nous avons utilisé des algorithmes génétiques (AG). L'évaluation de la solution est réalisée par une fonction multidimensionnelle mono-objectif sur la combinaison de cours de l'étudiant s et leur apport en termes de soft skills. Trois fonctions sont utilisées pour évaluer les combinaisons : 1) une différence linéaire entre le soft skill estimé et le seuil, 2) une fonction quadratique par morceaux à gauche et à droite des seuils et 3) une fonction par morceaux exponentielle (cf. Équation 2).

$$score_{is} = f(\widehat{SSkill}_i) = \begin{cases} -\epsilon^{1.5(Seuil_m - \widehat{SSkill}_i)} & \text{if } \widehat{SSkill}_i < Seuil_m \\ \epsilon^{\widehat{SSkill}_i - Seuil_m} & \text{if } \widehat{SSkill}_i \geq Seuil_m \end{cases} \quad (2)$$

3 Analyse et résultats préliminaires

Le paramétrage utilisé pour notre AG (X=0.66, M=1/16, $Nb_{executions} = 100$) a été empiriquement défini. Les jeux de données générés ont été divisés avec un ratio de 80% (entraînement)- 20% (test). Nos résultats se basent sur trois métriques : le pourcentage 1) des étudiants qui ont passé tous les seuils, 2) de ceux qui ont passé au moins un seuil mais pas tous et 3) d'aucun seuil validé. Le tableau 3 montre nos résultats préliminaires. L'on peut voir que toutes les fonctions proposées permettent d'augmenter (entre 0% et 11%) le pourcentage moyen d'étudiants qui réussissent les seuils dans les 4 spécialisations.

4 Discussion et conclusions

Notre AG a été implémenté en Python. Les expérimentations ont été conduites sur un processeur Intel$^{(R)}$ Core$^{(TM)}$ i7-10810U @ 1.10GHz et 32 Go de RAM.

TABLE 3 – Comparaisons métriques en moyenne entre les fonctions. Les trois pourcentages sont ordonnés de la manière suivante : (tout, partiellement et aucun réussi)

Fitness	Spéc. 1	Spéc. 2	Spéc. 3	Spéc. 4
Sans Rec.	(29%,39%,32%)	(26%,60%,14%)	(25%,62%,13%)	(24%,63%,13%)
Comp. L	(34%,34%,32%)	(30%,48%,22%)	(33%,46%,21%)	(35%,50%,15%)
Par. Comp. L	(31%,33%,36%)	(28%,52%,20%)	(28%,50%,22%)	(32%,53%,15%)
Comp. QR	(33%,35%,32%)	(33%,46%,21%)	(32%,49%,19%)	(37%,47%,16%)
Par. Comp. QR	(28%,33%,39%)	(26%,49%,25%)	(26%,51%,23%)	(29%,55%,16%)
Comp. E	(32%,34%,34%)	(30%,49%,20%)	(32%,48%,20%)	(35%,50%,15%)
Par. Comp. E	(31%,32%,37%)	(26%,54%,20%)	(27%,49%,24%)	(29%,54%,17%)

Le temps d'exécution d'une recommandation était de l'ordre de 15 secondes (soit de 50 minutes par chaque 200 étudiants). Nous avons donc parallélisé les calculs pour réduire les temps totaux d'exécution. Les résultats préliminaires montrent une polarisation de la population, avec toutefois des changements positifs (étudiants ayant réussi tous les seuils) plus importants. Néanmoins, les métriques utilisées pour le moment ne peuvent à elles seules déterminer quelle fonction de score est à préférer pour ce type de problème, car chacune d'entre elles repose sur une logique différente. En plus, il y a certaines considérations à prendre en compte, qui constituent nos perspectives. Premièrement, la complexité croissante du problème avec le nombre de cours et sa réalité terrain. Deuxièmement, les fonctions de fitness proposées fonctionnent avec des mesures continues de soft skills, et devraient être adaptées dans le cas des mesures catégoriques. Finalement, nos résultats soulèvent des questions éthiques à étudier.

Remerciement

Cette contribution a été réalisée dans le cadre du projet de recherche SUCCESS, financé par l'i-SITE ULNE.

Références

1. Almonte, R. : A Practical Guide to Soft Skills : Communication, Psychology, and Ethics for Your Professional Life. Routledge, Taylor & Francis Group (2021), https ://doi.org/10.4324/9781003212942
2. Barrie, S. : Understanding what we mean by generic attributes of graduates. Higher Education **51**, 215–241 (01 2006)
3. Chamorro-Premuzic, T., Arteche, A., Bremner, A., Greven, C., Furnham, A. : Soft skills in higher education : Importance and improvement ratings as a function of individual differences and academic performance. Educational Psychology **30**, 221–241 (03 2010)
4. Hartig, J., Höhler, J. : Multidimensional irt models for the assessment of competencies. Studies in Educational Evaluation **35**(2), 57–63 (2009). https://doi.org/https://doi.org/10.1016/j.stueduc.2009.10.002, assessment of Competencies
5. Lebis, A., Humeau, J., Fleury, A., Vermeulen, M. : Le cursus académique personnalisé dans une approche par compétences avec érosion : étude d'un nouveau problème fondamental. In : 10e Conférence sur les Environnements Informatiques pour l'Apprentissage Humain. pp. 94–105. Fribourg, Switzerland (2021)

Impact de la Réalité Virtuelle en formation à distance sur les conditions d'apprentissage

Laetitia Pluton[1], Erick Stattner[2]

[1] Université des Antilles, CRREF, 97139 Guadeloupe, France
laetitia.pluton@etu.univ-antilles.fr
[2] Université des Antilles, LAMIA, 97157 Guadeloupe, France
erick.stattner@univ-antilles.fr

Résumé. L'usage du numérique s'est fortement démocratisé dans les établissements d'enseignement, comme complément à la formation traditionnelle, puis massivement accentué par la crise sanitaire, durant laquelle on a vu se développer les fonctionnalités des outils et les pratiques des enseignants. Dans ce travail, nous étudions la pertinence d'avoir recours à la Réalité Virtuelle (RV) comme moyen plus immersif, pour maintenir des cours à distance favorisant les apprentissages. L'objectif est de mesurer les apports de la RV en formation en comparaison aux cours en ligne. Deux axes sont ainsi abordés : l'apport de fonctionnalités favorisant les apprentissages et la perception des utilisateurs à travers leur vécu. Les résultats obtenus mettent en lumière la pertinence de cette technologie pour la formation, ses limites, et les défis qu'elle devra relever.

Mots-clés : Réalité virtuelle, Apprentissages, Expérience utilisateur, Enseignement à distance, Metaverse

Abstract. The use of digital technology has been greatly democratized in schools, as a complement to traditional learning. This use was then massively accentuated by the health crisis, which has allowed to develop the functionalities of the tools and the practices of the teachers. In this work, we study the relevance of using a more immersive technology, the Virtual Reality (VR), to maintain distance courses that promote learning. The objective is to measure the contributions of VR in comparison to online courses. Two axes are thus addressed: the development of functionalities promoting learning and the perception of users through their experience. The results obtained highlight the relevance of this technology for learning, its limits, and the challenges it will have to meet to fully impose itself.

Keywords: Virtual reality, Learning, User experience, Distance education, Metaverse

1 Introduction

La crise sanitaire de 2020 a conduit à la fermeture régulière des établissements scolaires, universitaires, de formation et entraîné la mise en œuvre une continuité pédagogique. Les enseignants et les apprenants ont dû faire preuve d'une capacité d'adaptation face à une modalité d'enseignement et d'apprentissage souvent sous-utilisée et peu maîtrisée, ainsi qu'à une maîtrise minimale mais rapide des outils numériques à leur disposition [1]. Bien que récents, les travaux sur cette nouvelle modalité d'enseignement ont mis en exergue les difficultés vécues par les différents acteurs [1], [2]. Pour nous, il s'agit de mesurer en quoi l'utilisation de la RV en formation à distance permet de favoriser les apprentissages, en comparaison aux cours à distance synchrones tels que menés jusque-là. Ainsi, dans ce travail, nous faisons l'hypothèse que le recours à la Réalité Virtuelle (RV), comme alternative aux cours en ligne synchrones, permettrait de maintenir des cours à distance favorisant les apprentissages. Nous nous appuyons pour cela sur le vécu des étudiants et des enseignants.

La distance dans la relation pédagogique caractérise l'enseignement à distance dont le principal défi réside dans sa capacité à supprimer l'absence : celle du médiateur par rapport aux savoirs, celle des interrelations et interactions entre apprenants et avec les enseignants [3]. Pour qu'il y ait apprentissage, en présence ou à distance, l'enseignement doit proposer des situations qui comportent une construction ou une résolution de problème, au sens de Pastré et des chercheurs de la didactique professionnelle [4], [5]. Or depuis ces deux dernières années, bien que la crise sanitaire ait accéléré le développement d'outils technopédagogiques, le constat reste mitigé. Les récents travaux sur le déploiement des cours à distance ont révélé un impact négatif sur les apprentissages. La principale raison évoquée est l'absence d'actions et d'interactions entre participants [6]. Pour autant, la formation à distance tend à se développer, il apparaît donc important de s'y intéresser en tenant compte des limites de l'existant, d'où notre intérêt pour l'utilisation de la RV en formation.

La RV simule un environnement, modélisé en 3 dimensions. L'utilisateur, représenté par un avatar, y accède via un casque qui le projette dans un monde virtuel avec lequel il peut interagir. Expérimentée dans des dispositifs de formation nécessitant des plateaux techniques spécifiques ou onéreux, la RV semble proposer des alternatives intéressantes dans le cadre de simulations. Les auteurs de ces travaux s'accordent tous sur l'intérêt de la RV comme outil de simulation et d'immersion favorisant les apprentissages [7]. Pour autant, en l'état actuel de nos connaissances, il existe très peu de travaux relatifs à l'apport de l'utilisation de la RV utilisant les salles de classes virtuelles comme alternative aux cours en ligne synchrones.

2 Méthode

Pour réaliser cette étude, nous avons équipé de casques de RV une promotion composée de 24 étudiants en Master 1 et 2 en Méthodes Informatiques Appliquées à la Gestion de l'Entreprise (MIAGE), et de cinq enseignants volontaires intervenant dans les enseignements de ces deux promotions. Parmi ces 24 étudiants, 19 poursuivent le parcours depuis la Licence 3 et ont connu trois modalités de cours : en présentiel, en

ligne et l'expérimentation en RV, ce qui leur donne une vision globale du dispositif. Durant la pandémie, tous les cours ont intégralement été réalisés à partir de l'application *Zoom*, choisie par l'université. Les enseignements en RV se sont inscrits dans une période post-pandémique émaillée par des confinements partiels et des blocages routiers résultant de manifestations sociales. Cette modalité a été introduite dans six enseignements (3 en M1 et 3 en M2) ce qui représente un total de 23 séances de 2h dispensées en RV, réparties sur le semestre. Avant le premier cours en RV, chaque promotion a réalisé une séance de prise en main de l'outil de 2h (mise en service, installation et découverte des outils). Il est à préciser que sur les enseignements choisis, c'est une modalité hybride qui a été mise en place : cours en présentiel et cours en RV.

Nous avons utilisé l'application *Horizon Workrooms* qui permet de collaborer dans des salles de cours virtuelles pouvant impliquer jusqu'à une vingtaine de participants. L'environnement reproduit une salle de cours et permet aux participants de discuter, se déplacer et aller écrire au tableau.

Les données ont été recueillies à partir de questionnaires en ligne et d'entretiens semi-directifs individuels et collectifs. Le questionnaire et les entretiens menés avant de débuter l'expérimentation en RV avaient pour objectifs d'identifier des indicateurs de l'expérience utilisateur du cours à distance synchrone ; ceux administrés à la fin ont servi à identifier les indicateurs de l'expérience utilisateur de la RV.

Nous avons ainsi collecté 28 questionnaires avant expérimentation, et 26 questionnaires post expérimentation. Nous avons également réalisé 27 entretiens individuels avant expérimentation RV, 3 entretiens collectifs, post RV. Nous avons cherché, à travers ces indicateurs, les éléments nous permettant de vérifier l'apport dans les apprentissages d'un outil numérique plus immersif, et d'en identifier les adaptations possibles pour une amélioration de l'expérience utilisateur.

3 Résultats expérimentaux

Pour étudier la plus-value de la RV, nous avons tout d'abord choisi de nous intéresser aux fonctionnalités offertes par chacun des environnements de formation à distance. Nous avons cherché à identifier le niveau de maîtrise des fonctionnalités des outils des participants, leur niveau de satisfaction, ainsi que les avantages apportés par les fonctionnalités utilisées. Nous nous sommes ensuite intéressés au vécu des participants. Les données ont permis d'identifier un impact négatif du cours à distance synchrone sur l'attitude "en classe" et l'implication dans les activités proposées, ainsi que ses limites. On retrouve parmi les impacts les plus fréquemment évoqués : le manque d'interactions, les difficultés de concentration, les distractions offertes par leur environnement réel. Après expérimentation, l'analyse montre que l'immersion totale et les fonctionnalités offertes par la technologie ont renforcé les interactions et l'intérêt des étudiants. La liberté d'interactions et de prise de parole leur permettent d'être plus impliqués dans les activités et renforcent les échanges des participants.

Au regard des résultats, nous observons que les cours en RV semblent se dérouler dans un environnement qui réunit des conditions favorables aux apprentissages. Les recherches en didactique [4],[7] ont montré que la construction des apprentissages nécessite une situation qui (1) mettre en relation un formateur et un apprenant qui interagissent dans une relation de médiation, (2) contraigne l'apprenant à mobiliser ses

ressources antérieures, (3) un contexte, qui conditionne, contraint et autorise les conduites des apprenants dans une situation donnée et (4) un milieu, défini par les aspects matériels et la dimension sémiotique. Ces conditions amènent l'apprenant à construire de nouvelles réponses par adaptation des schèmes et dispositions.

En nous appuyant sur les travaux de Sensevy [4] et Pastré [5], nous pouvons observer à partir des éléments issus des questionnaires et des entretiens que les cours en RV réunissent davantage les conditions favorables aux apprentissages que ne le font les cours à distance synchrones. En recréant les principales caractéristiques d'un cours en présentiel, la compréhension est facilitée et l'intérêt suscité, en particulier grâce aux interactions entre participants. Le sentiment d'isolement est diminué, et les possibilités de distraction sont rendues quasi impossibles. En ce sens, l'utilisation de la RV en formation nous semble être prometteuse pour l'enseignement à distance bien que des améliorations techniques restent encore à mettre en œuvre.

4 Conclusion et perspectives

Dans ce travail, nous avons mené une expérience autour de l'usage de la réalité virtuelle et de son apport, comparé aux cours à distance synchrones. L'expérience menée a permis d'observer que l'usage d'environnements simulés par la RV permet aux apprenants de s'immerger dans des situations de formation plus satisfaisantes, interactives et impliquantes. Si la technologie s'avère donc prometteuse dans le développement de formations à distance, des défis techniques et fonctionnels sont encore à relever pour répondre pleinement aux exigences des utilisateurs. Dans nos travaux futurs, nous entendons poursuivre cette expérience avec d'autres types de formation et proposer des indicateurs supplémentaires pour améliorer notre compréhension de l'apport de la RV.

Références

1. Boudokhane-Lima, F., Felio, C., Lheureux, F., Kubiszewski, V.: L'enseignement à distance durant la crise sanitaire de la Covid-19 : le faire face des enseignants en période de confinement, RFSIC, nº 22, mai 2021, doi: 10.4000/rfsic.11109.
2. Issaieva, E., Odacre E., Lollia, M., Joseph-Theodore, M.: Enseigner et apprendre en situation de pandémie : caractéristiques et effets sur les enseignants et les élèves, Formation-profession, vol. 28, nº 4 hors-série, p. 1, déc. 2020, doi: 10.18162/fp.2020.702.
3. Jacquinot, G.: 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.
4. Schneuwly, B., Sensevy G., Mercier, A. (dir.): Agir ensemble : l'action didactique conjointe du professeur et des élèves, Revue française de pédagogie Recherches en éducation, nº 160, Art. nº 160, septembre. 2007, doi: 10.4000/rfp.906.
5. Pastré, P.: Situation d'apprentissage et conceptualisation, Recherches en éducation, nº 12, Art. nº 12, novembre 2011, doi: 10.4000/ree.5085.
6. Granjon, Y.: La perception de l'enseignement à distance par les étudiants en situation de confinement : premières données, Distances et Médiations des savoirs, nº 33, 2021, doi: https://doi.org/10.4000/dms.6166.
7. Harfouche, A. L., Nakhle, F.: Creating Bioethics Distance Learning Through Virtual Reality, Trends in Biotechnology, vol. 38, nº 11, p. 1187□1192, novembre 2020, doi: 10.1016/j.tibtech.2020.05.005.

Analyse d'une stratégie d'autorégulation : La recherche d'aide dans l'apprentissage de la programmation sur une plateforme d'apprentissage en ligne

Marine Roche[1], Fahima Djelil[2], Jérôme Eneau[1], Jean-Marie Gilliot[2], Geneviève Lameul[1] et Hugues Pentecouteau[1]

[1] CREAD, Université Rennes 2, Rennes, France
prenom.nom@univ-rennes2.fr
[2] LabSTICC, IMT Atlantique Bretagne-Pays de la Loire, Brest, France
fahima.djelil@imt-atlantique.fr, jm.gilliot@imt-atlantique.fr

Résumé. Cette communication est issue de la recherche ANR xCALE, qui a pour objet d'analyser la manière dont les apprenants s'autorégulent quand ils utilisent des plateformes d'apprentissage en ligne. Il s'agit plus particulièrement d'interroger une stratégie d'autorégulation : la recherche d'aide. Nous avons cherché à identifier et caractériser les différentes aides mobilisées en décrivant les stratégies d'autorégulation de recherche d'aide dans le cadre d'une formation pour adultes dans le domaine du numérique.

Mots-clés : Autorégulation, Help Seeking, Plateforme d'apprentissage en ligne, Formation d'adultes, Situations d'apprentissage.

Abstract. This paper comes from a research entitled xCALE which aims to analyze how learners self-regulate during an online training. We examine a self-regulated learning strategy: help seeking. The objective is to identify and describe the practices of learners asking for help during the use of online training and their self-regulated learning strategies.

Keywords: Self-Regulated Learning, Help Seeking, Online Training, Adult Education, Learning Situations.

1 Introduction

Cette contribution s'inscrit dans la recherche ANR xCALE (eXplaining Competency and Autonomy development in Learning Environments). La recherche associe des chercheurs en EIAH, en IA et en sciences de l'éducation et de la formation. L'objectif consiste à développer, expérimenter et évaluer un dispositif de formation en ligne qui propose des soutiens didactiques. Le but est de pouvoir proposer des interventions adaptées, parfois automatisées, afin de soutenir les apprentissages et l'autorégulation des

apprentissages. En effet, suivre une formation en ligne demande de la part des apprenants, notamment, de savoir gérer sa motivation, d'être autonome ou encore d'être en mesure d'organiser ses apprentissages [1]. La plupart des apprenants ne sont pas formés à suivre des formations en ligne et rencontrent des difficultés. Dans les travaux les plus récents, le rôle des stratégies d'autorégulation des apprentissages a été mis en évidence [2][3]. En effet, apprendre dans un environnement en ligne demande aux apprenants de réguler leurs apprentissages : prendre des décisions, vérifier les apprentissages réalisés, chercher de l'aide, etc. [4]. On peut ainsi s'interroger sur les stratégies d'autorégulation utilisées lors de formation en présentiel mobilisant des plateformes d'apprentissage en ligne. Dans ce papier, nous présenterons les résultats issus de la phase exploratoire des chercheurs en sciences de l'éducation et de la formation et nous nous centrerons sur une stratégie d'autorégulation : la recherche d'aide (*help seeking*).

2 L'autorégulation des apprentissages et la recherche d'aide

Zimmerman définit l'autorégulation comme « l'intensité avec laquelle l'individu est aux plans de la métacognition, de la motivation et de la conduite un participant actif dans ses processus d'apprentissages » [5]. Zimmerman et Pons [6] proposent quatorze stratégies d'autorégulation des apprentissages. Dans ce papier, nous allons nous intéresser à deux stratégies : *seeking information*, les efforts déployés par les apprenants pour obtenir des informations supplémentaires auprès de sources « non sociales », et *seeking social assistance*, les efforts déployés par les apprenants pour solliciter de l'aide auprès de leurs pairs, des enseignants et des adultes [6]. Ces deux stratégies d'autorégulation ont été choisies parce qu'il est possible de réaliser des recueils de données pour toutes les disciplines mobilisées dans la recherche (EIAH, IA et sciences de l'éducation et de la formation). Zimmerman et Pons distinguent la recherche d'aide auprès d'une personne et la recherche d'informations où il s'agit de rechercher des informations dans des ressources. Puustinen et Rouet [7] reviennent sur cette distinction et proposent un continuum où les différentes aides seraient différenciées selon leur capacité à adapter leurs réponses aux besoins des apprenants. En effet, pour ces auteurs, ce qui différencie un système d'information « passif » d'une interaction avec une personne est la prise en compte de ses caractéristiques et du contexte de la recherche d'aide. Ainsi, à gauche du continuum se situent les aides qui ne peuvent pas s'adapter aux apprenants et qui fournissent toujours la même réponse indépendamment du niveau ou des connaissances préalables de l'apprenant. Et à l'opposé, se situent les aides qui s'adaptent toujours aux besoins de l'apprenant.

On peut également souligner que la recherche d'aide en ligne est différente de la recherche d'aide en présentiel. En effet, les environnements en ligne possèdent des caractéristiques spécifiques comparées à la salle de classe : disponibilité des ressources, réflexion sur la question à formuler et rédaction, absence d'indice sur les caractéristiques sociales et scolaires, pas d'immédiateté dans les réponses, nature publique des messages sur les forums/chat, etc. [3][8][9]. Ainsi, on peut supposer que l'aide ne sera pas la même s'il s'agit d'une personne présente dans la salle pour aider ou s'il s'agit de mettre à disposition des ressources en ligne. On peut ainsi interroger les pratiques d'aide des apprenants d'une formation en présentiel mobilisant des ressources en ligne.

3 Terrain de la recherche et méthodologie

Le terrain de la recherche est une formation destinée aux adultes et vise une insertion professionnelle dans les métiers du numérique. La formation mobilise des plateformes d'apprentissage en ligne développées par France IOI dont le responsable participe à la recherche. Le recueil de traces informatiques est ainsi simplifié et le projet peut également permettre de développer un outil d'aide sur la plateforme. Le module choisi pour la recherche concerne l'apprentissage du code informatique et de l'algorithmique. D'une durée de 40 heures, il est intitulé « De l'algorithmique au code ». Durant ces séances, les stagiaires travaillent en autonomie sur la plateforme en ligne dans la même salle et en présence des formateurs. La plateforme d'apprentissage en ligne utilisée, Algorea Serious Game, vise l'apprentissage de l'algorithmique par la pratique en résolvant des exercices ludiques (appelés « puzzle »). Il s'agit par exemple, dans un espace virtuel, de programmer les déplacements d'un robot dans un parcours afin qu'il plante des fleurs. Dans les exercices, les instructions disponibles ainsi que leur nombre sont limitées, il existe quatre niveaux de difficulté par puzzle et les exercices peuvent être réalisés en codant avec des blocs ou en Python. Nous nous sommes demandés quelles sont les aides mobilisées par les stagiaires et quelles sont leurs caractéristiques ?

23 entretiens semi-directifs ont été recueillis auprès des stagiaires volontaires de la formation. 8 femmes et 15 hommes âgés entre 17 et 58 ans ont répondu à nos questions. Le guide d'entretien est composé de cinq parties : le profil des stagiaires ; les raisons pour lesquelles ils sont inscrits et leur projet professionnel ; l'environnement de travail ; l'environnement social et l'utilisation de la plateforme d'apprentissage en ligne pour apprendre la programmation et l'algorithmique.

4 Résultats

À partir des entretiens et des observations des séances, nous pouvons dresser une liste des aides disponibles pour les apprenants :

- sur la plateforme : l'aide fournit des informations sur les instructions (à noter qu'il n'existe pas de forum au moment de la rédaction de ce papier) ;
- l'usage d'internet pour trouver de l'aide : recherche libre (à noter que ChatGPT n'était pas connu du grand public au moment du recueil de données) ;
- la sollicitation des autres stagiaires : l'entraide ;
- la sollicitation des formateurs.

À partir des propos des stagiaires et de la littérature, nous avons cherché à caractériser les différentes aides mobilisées. Cinq dimensions sont retenues :

- la qualité des informations obtenues : la fiabilité de l'aide, l'expertise ou la compétence de la personne, la capacité à fournir des informations de qualité ;
- l'accessibilité de l'aide : sa capacité à être accessible à tous, le fait que l'aide soit facilement accessible à un stagiaire et qu'il est aisé de l'interroger ;
- la personnalisation de la réponse : la capacité de l'aide à tenir compte des caractéristiques de l'apprenant (comme le niveau scolaire) ;

- le type de demande : permet de savoir si la demande vient d'une initiative du stagiaire et si la demande s'adresse à une autre personne ;
- le type d'aide fournie : permet de savoir s'il s'agit d'une information, de la solution ou si elle permet une réflexion sur le problème rencontré.

5 Conclusion

Dans cette contribution, nous avons présenté et caractérisé les différents types d'aides mobilisées par les stagiaires. ChatGPT ne faisait pas partie des aides existantes, il serait aujourd'hui intéressant d'analyser son usage dans la recherche d'aide grâce aux dimensions présentées. Enfin, nous souhaitons revenir sur l'aspect pluridisciplinaire de la recherche et les apports entre les différentes disciplines. Les résultats présentés ici sont issus de la phase exploratoire d'un recueil de données en sciences de l'éducation et de la formation. Les apports pour la recherche sont multiples : rendre visible ce qui ne l'est pas dans les traces d'apprentissages ; définir les stratégies d'autorégulation des apprenants utilisant la plateforme d'apprentissage en ligne ; caractériser les aides des apprenants. Ces éléments pourront être pris en compte dans l'accompagnement ou les recommandations de l'outil d'aide développé sur la plateforme d'apprentissage en ligne.

Remerciements : Ce travail est financé par le projet ANR xCALE (ANR-20-CE38-0010).

Références

1. Molinari, G., Poellhuber, B., Heutte, J., Lavoué, E., Sutter Widmer, D., Caron, P.-A.: L'engagement et la persistance dans les dispositifs de formation en ligne : regards croisés. Distances et médiations des savoirs 13 (2016).
2. Greene, J., Moos, D., Azevedo, R.: Self-regulation of learning with computer-based learning environments. New directions for teaching and learning 126, 107-115 (2011).
3. Kizilcec, R. F., Pérez-Sanagustin, M., Maldonado, J. J.: Self-regulated learning strategies predict learner behavior and goal attainment in Massive Open Online Courses. Computers & education 104, 18-33 (2017).
4. Azevedo, R., Cromley, J.: Does training on self-regulated learning facilitate students' learning with hypermedia?. Journal of educational psychology 96(3), 523-535 (2004).
5. Zimmerman, B. J.: A social cognitive view of self-regulated academic learning. Journal of educational psychology 81(3), 329-339 (1989).
6. Zimmerman, B. J., Pons, M.: Development of a structured interview for assessing student use of self-regulated learning strategies. American educational research journal 23(4), 614-628 (1986).
7. Puustinen, M., Rouet, J. F.: Learning with new technologies: Help seeking and information searching revisited. Computers & Education 53(4), 1014-1019 (2009).
8. Koc, S., Liu, X.: An Investigation of Graduate Students' Help-Seeking Experiences, Preferences and Attitudes in Online Learning. Turkish Online Journal of Educational Technology 15(3), 27-38 (2016).
9. Mahasneh, R. A., Sowan, A. K., Nassar, Y. H.: Academic help-seeking in online and face-to-face learning environments. E-Learning and Digital Media 9(2), 196-210 (2012).

Un dispositif de suivi des apprentissages pour soutenir l'orchestration de la classe

Madjid Sadallah, Maria Teresa Segarra, Jean-Marie Gilliot et Issam Rebai

IMT Atlantique, Lab-STICC UMR CNRS 6285, F-29238 Brest, France
madjid.sadallah@imt-atlantique.fr, mt.segarra@imt-atlantique.fr,
jm.gilliot@imt-atlantique.fr, issam.rebai@imt-atlantique.fr;

Résumé. Le concept d'orchestration de la classe vise à saisir la complexité associée à la gestion des activités d'apprentissage synchrones (où apprenants et enseignants se retrouvent pour réaliser les activités). Notre objectif est de construire un dispositif permettant d'assister l'enseignant dans ses tâches d'orchestration, en explorant les notions de scénarisation, de suivi et de rétroaction. Dans cet article, nous présentons SAVAStudio, un outil permettant à l'enseignant de planifier le déroulement d'activités synchrones et de récupérer les traces du déroulement réel de celles-ci.

Mots-clé : Orchestration · Scénarisation · Suivi · Classe hybride

Abstract. Classroom orchestration aims at capturing the complexity associated with managing synchronous learning activities (where learners and teachers work together to complete activities). Our goal is to build a tool to support the teacher in his orchestration tasks, by exploring the notions of scripting, monitoring and feedback. In this article, we present SAVAStudio, a tool that allows the teacher to plan the conduct of synchronous activities and to retrieve traces of their actual execution.

Keywords: Orchestration · Scripting · Monitoring · Blended classroom

1 Introduction

Planifier le déroulement des activités d'apprentissage et l'adapter en fonction de l'avancement en séance constitue un enjeu particulièrement difficile pour l'enseignant. La notion d'*orchestration* apporte des concepts pertinents pour encadrer les défis auxquels les enseignants sont confrontés en classe, de la planification à la gestion du déroulement [2]. L'orchestration revient à gérer les différentes activités se déroulant dans différents contextes éducatifs et à différentes granularités (individu, groupe, classe, institution), en utilisant des ressources et outils de manière synergique [7]. L'un des défis de l'orchestration est centré sur la planification des sessions et leur adaptation à la volée en fonction des comportements et des besoins des apprenants [1]. Cependant, la capture rapide et le traitement à la volée des informations de la classe pour faciliter la coordination des activités d'apprentissage et identifier les problèmes potentiels sont difficiles car ils exigent

une répartition constante de l'attention des enseignants à différents niveaux [6]. Pour mieux soutenir les enseignants, des outils communément appelés "*technologie d'orchestration*" peuvent être utilisés en classe pour capturer, analyser et visualiser la manière dont les apprenants progressent dans leurs activités [4] et ainsi adapter à la volée celles-ci si nécessaire.

Un problème majeur de ces technologies est leur manque de prise en compte de la notion de *charge d'orchestration* des enseignants, c.-à-d., la charge cognitive qu'ils encourent lorsqu'ils régulent à la volée de multiples activités et processus d'apprentissage [7]. Cette charge ajoute de la complexité qui exige la prise de mesures ou de décisions instantanément [7]. En outre, certains outils, en plus d'être intrusifs, tentent de proposer un soutien automatique excessif, susceptible d'aller à l'encontre de l'autonomie et du sens des responsabilités de l'enseignant [5]. Il est par conséquent nécessaire de s'orienter vers des approches de conception centrée sur l'humain [5, 8], qui facilitent l'incorporation des principales parties prenantes, notamment les enseignants, dans la conception et la mise en œuvre des technologies d'orchestration [3]. Notre objectif est de construire un outil pour les enseignants basé sur l'analyse de traces récupérées pendant les séances d'apprentissage, afin de faciliter la gestion de la classe en adoptant une approche participative avec les enseignants.

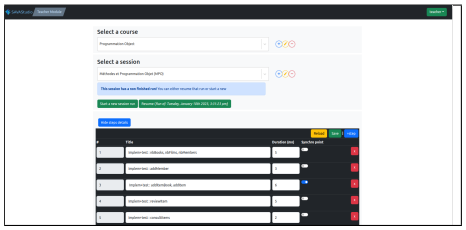
2 Un outil de suivi et d'orchestration synchrone

Approche méthodologique. Parce que notre objectif est d'impliquer les enseignants en situation d'usage réel, nous adoptons une approche DBR (*Design Based Research*) [9]. Nous suivons les principes de conception d'outils d'orchestration de Dillenbourg [2]: (1) *Contrôle* (permettant à l'enseignant de superviser la classe), (2) *Visibilité* (assurant la transparence des actions de tous les acteurs), (3) *Flexibilité* (permettant une adaptation facile aux changements), (4) *Physicalité* (intégrant l'outil à l'environnement physique) et (5) *Minimalisme* (fournissant uniquement les fonctionnalités nécessaires). D'un point de vue méthodologique, nous avons adopté une approche participative en organisant des ateliers de co-conception avec 7 enseignants sur 5 mois. Nous avons créé des maquettes, puis un prototype fonctionnel qui a été testé avec un petit groupe d'enseignants avant d'être amélioré et testé avec deux classes d'environ 20 apprenants

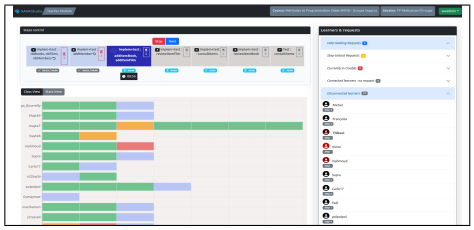
L'outil SAVAStudio. Le résultats des ateliers est SAVAStudio[1], une application dédiée au suivi et à l'orchestration d'une classe hybride (Figure 1). Elle est constituée d'un module *enseignant* et d'un module *apprenant*.

Le module enseignant est dédié à 1) la scénarisation et 2) au suivi de la classe, notamment pour les activités de mise en situation. L'enseignant peut définir des scénarios de déroulement d'une activité sous la forme d'un enchaînement d'étapes (Figure 1a). Chaque étape peut avoir une durée nominative, et peut être bloquante, et dans ce cas, un apprenant ne peut aller au-delà de l'étape

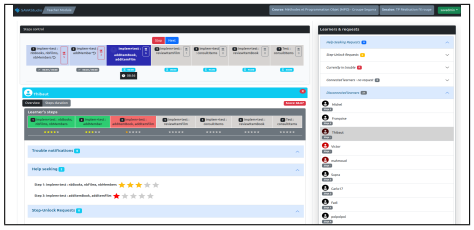
[1] Une instance de l'application est accessible à l'adresse : https://savastudio.enstb.org/



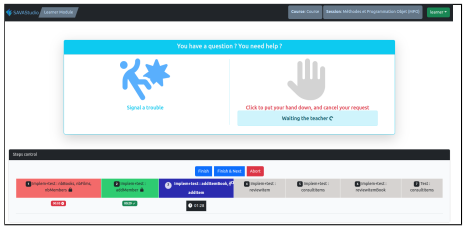
(a) Interface de scénarisation



(b) Suivi de l'avancement de la classe



(c) Suivi de l'avancement d'un apprenant



(d) Module d'interaction de l'apprenant

Fig. 1: Interfaces enseignant et apprenant de SAVAStudio

jusqu'à ce que l'enseignant la déverrouille. Ce mécanisme de synchronisation des apprenants peut être utilisé par l'enseignant afin de faire un point global à la classe sur un aspect précis de l'activité.

Pour le suivi de la classe, l'orchestration est faite en suivant le scénario défini par l'enseignant. Ce scénario et son rythme de progression prescrit sont utilisés par l'outil pour détecter les apprenants ayant des difficultés à progresser. Aucune décision de remédiation n'est prise à la place de l'enseignant : l'outil lui permet d'observer la progression globale de la classe (quel apprenant se trouve à quelle étape) (Figure 1b) afin qu'il puisse déterminer les actions à entreprendre sur la base des éléments présentés. Il lui offre aussi la possibilité de voir les détails de chaque apprenant, en affichant l'étape en cours, l'historique et le résultat des étapes précédentes, ses différentes requêtes (actuelles et passées), notamment si pour une étape donnée l'apprenant l'a terminé dans le temps prévu ou a abandonné les étapes précédentes et s'il sollicite de l'aide (Figure 1c). La *criticité* des demandes correspondant au temps d'attente. Un score global de niveau de compréhension basé sur son historique est associé à chaque apprenant.

Le module apprenant (Figure 1d) permet à ce dernier de déclarer son avancement dans les étapes et de signaler son niveau de compréhension à chaque fin d'étape, et aussi de solliciter l'enseignant, via des boutons de demande d'aide et de signalisation de problèmes (type *panic button*).

3 Conclusion et perspectives

À travers l'outil SAVAStudio, nous proposons d'explorer de nouvelles modalités d'interaction pour mieux assister l'enseignant dans l'orchestration de la classe.

Pour ce faire, nous avons suivi une approche impliquant les enseignants durant les phases de conception. L'outil proposé est l'un des rares travaux qui s'intéressent à capter l'interactivité entre apprenants et enseignants en séance synchrone et en environnement hybride. D'autre part, cette capture d'interaction est faite de deux manières : (1) *déclarative* : les apprenants indiquent leur avancement dans les étapes proposées par l'enseignant ainsi que leurs demandes d'aide ; (2) *analytique* : par le calcul des niveaux de compréhension et de difficulté des apprenants. Ces informations sont présentées à l'enseignant qui reste maître dans le choix des actions de remédiation. Au travers de cet outil, nous aspirons à démontrer qu'il est possible de proposer des retours utilisables par l'enseignant pendant son travail en classe. Il est actuellement en cours de test dans des classes pilotes. Cette première expérimentation permettra de mettre en place un premier protocole de mesure d'impact. Comme perspectives, nous souhaitons augmenter le moteur de suivi des apprenants par la définition de métriques plus élaborées, permettant de mieux caractériser leurs difficultés. Pour cela, nous nous penchons sur des méthodes d'intelligence artificielle appliquées en éducation. Dans un second temps, nous comptons élaborer un module d'assistance à l'enseignant ayant la capacité de proposer des actions d'orchestration, à base de l'état de la classe et des apprenants.

References

1. Amarasinghe, I., Hernandez-Leo, D., Michos, K., Vujovic, M.: An actionable orchestration dashboard to enhance collaboration in the classroom. IEEE Transactions on Learning Technologies **13**(4), 662–675 (2020)
2. Dillenbourg, P.: Design for classroom orchestration. Computers & education **69**, 485–492 (2013)
3. Dimitriadis, Y., Martínez-Maldonado, R., Wiley, K.: Human-centered design principles for actionable learning analytics. In: Research on E-Learning and ICT in Education, pp. 277–296. Springer (2021)
4. Faucon, L., Olsen, J.K., Haklev, S., Dillenbourg, P.: Real-time prediction of students' activity progress and completion rates. Journal of Learning Analytics **7**(2), 18–44 (2020)
5. van Leeuwen, A., Amarasinghe, I., Dimitriadis, Y., Martínez-Monés, A., Hernández-Leo, D., Hoppe, H., Wiley, K., Martinez-Maldonado, R., et al.: Teacher orchestration load: What is it and how can we lower the burden? In: ISLS Annual Meeting (2022)
6. van Leeuwen, A., Rummel, N.: Orchestration tools to support the teacher during student collaboration: a review. Unterrichtswissenschaft **47**(2), 143–158 (2019)
7. Prieto, L.P., Sharma, K., Kidzinski, Ł., Dillenbourg, P.: Orchestration load indicators and patterns: In-the-wild studies using mobile eye-tracking. IEEE Transactions on Learning Technologies **11**(2), 216–229 (2017)
8. Sadallah, M., Gilliot, J.M., Iksal, S., Quelennec, K., Vermeulen, M., Neyssensas, L., Aubert, O., Venant, R.: Designing lads that promote sensemaking: A participatory tool. In: Educating for a New Future: Making Sense of Technology-Enhanced Learning Adoption. pp. 587–593. Springer International Publishing, Cham (2022)
9. Wang, F., Hannafin, M.J.: Design-based research and technology-enhanced learning environments. Educational Technology Research and Development **53**(4), 5–23 (2005)

De quelques facteurs influençant le sentiment de compétences numériques des enseignant-es en début de carrière

Denise Sutter Widmer[1], Cécile Vassaux[2], Hana Bida[2], Mireille Bétrancourt[1], Stéphanie Boéchat-Heer[2]

[1] Université de Genève, TECFA, 1227 Carouge, Suisse
prenom.nom@unige.ch
[2] Haute École Pédagogique BEJUNE, 2503 Bienne, Suisse
prenom.nom@hep-bejune.ch

Résumé. Dans un contexte de transformation numérique de l'éducation, le développement des compétences numériques des enseignant-es débutant-e-s et leur intégration à la pratique éducative deviennent deux enjeux primordiaux. Dans ce cadre, l'étude menée par le projet ESCOFE cherche à comprendre comment évolue le sentiment de compétence numérique des enseignant-es débutant-e-s de la fin de leur formation d'enseignement aux premières années d'insertion professionnelle. Via l'administration d'un questionnaire à des enseignant-es en 1re et 3e année de pratique, elle cherche à identifier les obstacles et leviers pouvant influencer l'intégration du numérique au cours du temps et s'intéresse à deux éléments rarement pris en considération, le stress ressenti par les enseignant-es débutant-e-s face au numérique et les stratégies d'ajustement mobilisées afin de remédier aux difficultés rencontrées avec le numérique.

Mots-clés : Compétences numériques, sentiment d'auto-efficacité, stratégies d'ajustement, pratiques, formation.

Abstract. In a context of digital transformation of education, the development of digital competence of beginning teachers and their integration into educational practice become two key issues. In this context, the study conducted by the ESCOFE project seeks to understand how the feeling of digital competence of beginning teachers evolves from the end of their teacher training to the first years of professional insertion. A questionnaire was administered to teachers in their first or their third year of practice, to identify the obstacles and levers that can influence the integration of digital technology over time. Innovative elements of this study include the stress felt by beginning teachers as well as the adjustment strategies used to remedy the difficulties encountered with digital technology.

Keywords: Digital competence, feeling of competence, adjustment strategies, practices, training.

1 Contexte et enjeux

La formation initiale et continue des enseignant-es aux compétences numériques est une préoccupation européenne depuis près de 20 ans. Cependant, malgré les formations suivies dans le domaine, les enseignant-es débutant-e-s intègrent toujours peu voire pas du tout les technologies numériques dans leur pratique pédagogique [1] [2]. Les étudiant-es en formation en enseignement ne se sentent d'ailleurs pas plus formé-es aux compétences numériques en 2021-22 qu'avant la période de confinement [3].

De nombreux obstacles, institutionnels, organisationnels et individuels, expliquent la difficulté à développer un sentiment de compétence numérique et à intégrer les outils numériques dans les pratiques scolaires. Des facteurs tels que des réticences à l'utilisation de la technologie, notamment si la valeur ajoutée ne semble pas évidente, sont reconnus [4] comme frein à l'intégration du numérique. A contrario, les compétences numériques, un sentiment de compétence élevé [5], la collaboration et le soutien au sein de l'établissement par les pairs [6] et par la direction [7], la formation continue ainsi qu'une formation initiale englobant un apprentissage vicariant [8], sont des facteurs susceptibles de favoriser l'intégration du numérique. Enfin, les stratégies d'ajustement face aux difficultés pouvant être rencontrées lors de l'intégration d'outils numériques engendrant des situations d'enseignement nouvelles, génératrices de stress, peuvent être plus ou moins efficaces et faciliter leur utilisation ou, au contraire, constituer des freins à une intégration sereine du numérique [9].

Une récente méta-analyse [10] conduite dans le cadre du modèle TAM (Technology acceptance model) réunissant les résultats de 114 études a confirmé l'importance de ces variables sur l'intention d'usage et in fine sur l'utilisation effective des technologies numériques. Alors que l'effet de ces variables est bien établi, il y a peu de littérature sur leur interdépendance et leur évolution au cours du temps [11]. En particulier, il manque des études qui nous renseignent sur l'évolution du sentiment de compétences numériques ainsi que sur le transfert des compétences [12] de la formation initiale à l'insertion professionnelle en fonction de l'expérience effective de ces technologies dans le cadre de leur pratique.

2 Problématique et questions de recherche

Partant de ces différents constats, le projet ESCOFE a pour objectif d'étudier l'évolution du sentiment de compétence des enseignant-es débutant-es au fil des quatre premières années d'insertion professionnelle. Trois questions de recherche sont posées : 1) Dans quelle mesure les dispositifs de formation exercent-ils une influence sur le sentiment de compétence numérique et le niveau d'intégration du numérique en classe ? 2) Quels sont les facteurs pouvant constituer un obstacle ou un levier lors de l'intégration du numérique à l'école ? 3) A quels problèmes les individus sont-ils confrontés lorsqu'ils essaient d'appliquer les compétences acquises en formation initiale et quelles stratégies d'ajustement sont mises en place pour y remédier ? La gestion du stress par les enseignant-es qui utilisent le numérique et leurs stratégies d'ajustement méritent d'être étudiées comme nouveaux facteurs permettant de

comprendre l'évolution du sentiment de compétence des enseignants et des usages mis en place.

3 Méthode

Le projet ESCOFE lancé en avril 2022 vise à répondre à ces questions en suivant une méthode mixte incluant un volet quantitatif (questionnaire) et un volet qualitatif (focus groups et entretiens individuels) qui ne sera pas développé dans cette contribution. Construit sur la base de questionnaires existants (notamment la « Teachers'Sense of Efficacy Scale » (TSES) [13] et le « Brief COPE inventory » [14] pour les stratégies d'ajustement), le questionnaire du volet quantitatif a été administré à l'automne 2022 auprès de deux cohortes d'enseignant-es issu-e-s de cinq cantons romands (Genève, Vaud, Neuchâtel, Berne francophone, Jura) des degrés primaire et secondaire I (âge 12-15 ans). La première cohorte est composée de personnes entamant leur 1re année d'enseignement (diplômée en 2022), tandis que la seconde débute sa 3e année d'enseignement (diplômée en 2020).

Le sentiment de compétence numérique en début de carrière et l'impact de la formation sur celui-ci a été cerné par des questions portant sur les compétences numériques acquises pendant la formation initiale (par ex. « Connaître les démarches pédagogiques qu'implique l'usage du numérique »), les modules de cours ayant permis en formation de développer des compétences numériques, le sentiment d'efficacité lors de l'utilisation du numérique en classe (par ex. « Je suis capable de gérer la classe »), les stratégies d'ajustement mises en place face aux difficultés rencontrées avec le numérique (par ex. « Je mène des actions concrètes pour améliorer la situation »), les représentations du numérique et les usages numériques réalisés ou envisagés dans le futur (31 items d'usages possibles). Des questions spécifiques à chaque cohorte ont été intégrées au questionnaire : sur l'expérience de stage en lien avec le type de soutien provenant du formateur de terrain pour la cohorte 1 ; sur le degré d'intégration numérique (via des vignettes de situation) et les caractéristiques de l'établissement scolaire (soutien proposé, culture d'établissement, matériel à disposition) pour la cohorte 2. Les variables socio-démographiques (âge, sexe, établissement de formation, diplôme) sont abordées en début de questionnaire. Un deuxième questionnaire sera soumis aux deux cohortes une année après la première passation afin d'analyser l'évolution du sentiment de compétence numérique au fil de l'insertion professionnelle.

Des analyses descriptives seront effectuées pour identifier les obstacles et leviers à l'intégration du numérique. Des analyses de régressions simples et multiples sont prévues afin de déterminer dans quelle mesure les dispositifs de formation, les représentations du numérique influencent le sentiment de compétence des enseignant-es débutant-es, leur usage du numérique et le niveau d'intégration en classe. Un modèle corrélationnel permettra de vérifier si le type de stratégies d'ajustement mobilisé en cas de difficultés influence le sentiment de compétence de l'enseignant-e et si ces stratégies sont susceptibles d'évoluer au cours du temps.

4 Conclusion

Les résultats obtenus permettront de mieux comprendre les processus à l'œuvre dans l'évolution de la perception des enseignant-es des compétences numériques acquises durant la formation, et d'évaluer l'influence des stratégies d'ajustement et du soutien au sein des établissements. L'originalité de cette recherche est de se situer dans des cantons dont les formations et les politiques d'introduction de l'éducation numérique diffèrent, ce qui permettra de faire émerger des invariants et des facteurs variables. Les résultats devraient conduire à la formulation de recommandations aux institutions de formation pour améliorer les dispositifs de formation et aux directions d'école pour favoriser l'accompagnement et le soutien au sein des établissements scolaires.

Références

1. Reisoğlu, İ., Çebi, A.: How can the digital competences of pre-service teachers be developed? Examining a case study through the lens of DigComp and DigCompEdu. Computers and Education, 156 (2020).
2. Tondeur, J., Van Braak, J., Siddiq, F., Scherer, R.: Time for a new approach to prepare future teachers for educational technology use: Its meaning and measurement. Computers and Education, 94, 134–150 (2016).
3. Boéchat-Heer, S., Moor, M.: Perceptions des étudiant · e · s des compétences, représentations et usages numériques en formation à l'enseignement avant et après le confinement. In 6e colloque de l'AUPTIC•Education, pp. 2020–2022. Namur (2022).
4. Kretschmann, R.: Physical education teachers' subjective theories about integrating information and communication technology (ict) into physical education. TOJET: The Turkish Online Journal of Educational Technology (vol.14) (2015).
5. Boéchat-Heer, S.: L'adaptation des enseignants aux usages des MITIC. Sentiment d'auto-efficacité, formation et pratiques en classe. Thèse. Université de Fribourg (2009).
6. Starkey, L.: Supporting the digitally able beginning teacher. Teaching and Teacher Education, 26(7), 1429–1438 (2010).
7. IsaBelle, C., Lapointe, C., Chiasson, M.: Pour une intégration réussie des TIC à l'école : De la formation des directions à la formation des maîtres. Revue des Sciences de l'éducation, 28(2), 325–343 (2002).
8. Bandura, A.: Self-efficacy: The exercise of control. New York: Freeman. (1997).
9. Boéchat-Heer, S., González-Martínez, E.: Les difficultés et les stratégies d'ajustement des enseignants face à l'innovation technologique : trois études de cas. Sticef, 1 (2021).
10. Scherer, R., Siddiq, F., Tondeur, J.: The technology acceptance model (TAM): A meta-analytic structural equation modeling approach to explaining teachers' adoption of digital technology in education. Computers and Education, 128, 13–35 (2019).
11. Xie, K., Nelson, M. J., Cheng, S. L., Jiang, Z.: Examining changes in teachers' perceptions of external and internal barriers in their integration of educational digital resources in K-12 classrooms. Journal of Research on Technology in Education (2021).
12. Payn, M., Derrac, L., Alvarez, L.: Le transfert des compétences d'usages numériques. Interfaces, 11(3) (2022).
13. Tschannen-Moran, M., Woolfolk Hoy, A.: Teacher efficacy: Capturing an elusive construct. Teaching and Teacher Education, 17, 783-805 (2001).
14. Carver, C. S.: You want to measure coping but your protocol is too long: Consider the brief cope. International journal of behavioral medicine, 4(1), 92-100 (1997).

Index des auteurs

A

Ah-Tchine Emeline, 217
Ait Lahcen Ayoub, 324, 340
Alamargot Denis, 336
Alvarez Lionel, 271
Amadieu Franck, 312
Augereau Olivier, 24, 48

B

Badier Anaëlle, 232
Beffara Emmanuel, 308
Bétrancourt Mireille, 360
Bida Hana, 360
Bodin Martin, 308
Boéchat-Heer Stéphanie, 360
Bouchet François, 256, 295, 336
Bouton Maxime, 133
Broisin Julien, 36, 312
Bruel Jean-Michel, 68
Brumeaud Quentin, 271

C

Carron Thibault, 256, 316
Crespin Benoit, 223
Crétin-Pirolli Raphaëlle, 328, 332

D

De Lièvre Bruno, 42
Dechaux Eva, 166
Delaporte Patrick, 145
Deleau Romain, 145
Devillers Frédéric, 24
Devillers Laurence, 48
Dham Cedric, 244
Didier Jean-Yves, 112
Djelil Fahima, 244, 352
Douady Julien, 320

E

El Bouzekri El Idrissi Younès, 324, 340
El Kechaï Hassina, 205
Eneau Jérôme, 352
Ez-Zaouia Mohamed, 100

F

Fleury Anthony, 344
Folcher Viviane, 217

G

Ganier Franck, 24
Geoffre Thierry, 271
George Sébastien, 88, 121, 283
Gilliot Jean-Marie, 56, 352, 356
Gobard Anne-Claire, 166
Gril Albane, 283
Guigon Gaëlle, 316

H

Haddouche Anis, 244
Hattab Lucas, 121
Hervot Paul, 223
Hoffmann Christian, 244, 320
Huguenin Audrey, 80

I

Ihichr Adel, 324, 340

J

Jaccard Dominique, 80
Jaouadi Mariem, 80
Jolivet Sébastien, 166

K

Karami Abir, 344
Karoui Aous, 271
Kobylyanskaya Sofiya, 48

L

Laforcade Pierre, 88
Lagasse Marine, 42
Lallé Sébastien, 295
Lameul Geneviève, 352
Laurent Maxence, 80
Lavoué Elise, 193
Le Bigot Nathalie, 24
Lebis Alexis, 145, 344
Lefevre Marie, 232
Lefort Mathieu, 232
Lemoine Bérénice, 88
Leonard Marielle, 133
Lescure Emmanuel, 36
Levieux Sébastien, 24
Loiseau Mathieu, 193
Loison Valériane, 328
Loizon Anaïs, 217
Lopez Serena, 328
Loup Guillaume, 112, 262
Luengo Vanda, 295

M

Mandran Nadine, 80, 244, 308, 320
Marfisi-Schottman Iza, 100, 121
May Madeth, 283
Mercier Cendrine, 100
Michel Christine, 154
Minotti Kelly, 262
Molinier Rémi, 308
Monnier Sandra, 80
Muratet Mathieu, 178, 316

N

Ndiaye Mamoudou, 332
Nebel Léo, 336

O

Oliver-Quelennec Katia, 256
Ollivier Denis, 36
Otmane Samir, 262
Oustous Omar, 324, 340

P

Paggetti Marion, 217
Pelay Nicolas, 121
Pentecouteau Hugues, 352
Petelin Tanja, 205
Peter Yvan, 133
Pierrot Laëtitia, 154
Pinçon Claire, 256
Pinos Ullauri Luis Alberto, 344
Pinot Rémy, 145
Pirolli Fabrice, 328, 332
Pluton Laetitia, 348
Pons Mika, 68
Prior Estelle, 80

R

Raclet Jean-Baptiste, 36, 68
Raison Anaïs, 24
Rebai Issam, 356
Renault Valérie, 283
Roche Marine, 352

S

Sadallah Madjid, 56, 356
Sanchez Eric, 80
Schmitt Daniel, 42
Segarra Maria Teresa, 356
Silvestre Franck, 36, 68
Simonnet Enzo, 193
Stattner Erick, 348
Sutter Widmer Denise, 360

T

Touel Sofiane, 121

V

Van Den Noortgate Wim, 344
Vasilescu Ioana, 48
Vassaux Cécile, 360
Venant Rémi, 312
Verger Mélina, 295
Vermeulen Mathieu, 145, 316, 344

W

Wang Patrick, 166

X

Xuan Hien Mai Daniel, 112

EIAH2023 : 11ème Conférence sur les Environnements Informatiques pour l'Apprentissage Humain

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

12-16 juin 2023 Brest (France)

Actes de la onzième 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 Sébastien Iksal, Chrysta Pélissier,
Jean-Marie Gilliot et Patricia Marzin-Janvier

Du 13 au 16 juin 2023
IMT Atlantique, Brest
France

Nous remercions nos partenaires institutionnels :

ISBN n°978-2-901384-07-6