

Sticef

*Sciences et technologies de l'information et de la communication
pour l'éducation et la formation*

Volume 31, numéro 1, 2024

numéro spécial

**Sélection
de la Conférence EIAH 2023**

*sous la direction de
Chrysta PÉLISSIER,
Sébastien IKSAL,
Jean-Marie GILLIOT,
Patricia MARZIN-JANVIER*



DOI: 10.23709/sticef.31.1

Sticef

Volume 31
numéro 1, 2024

numéro spécial
Sélection de la
Conférence
EIAH 2023

© ATIEF, 2016

ISBN 978-2-901384-09-0

DOI: 10.23709/sticef.31.1 en ligne sur www.sticef.org

Le Code de la propriété intellectuelle n'autorisant, aux termes des paragraphes 2 et 3 de l'article L. 122-5, d'une part, que les « *copies et reproductions strictement réservées à l'usage privé du copiste et non destinées à une utilisation collective* » et, d'autre part, sous réserve de mention du nom de l'auteur et de la source, que « *les analyses et les courtes citations justifiées par le caractère critique, polémique, pédagogique, scientifique ou d'information* », « *toute représentation ou reproduction totale ou partielle faite sans le consentement de l'auteur, ou de ses ayants droit ou ayants cause, est illicite* » (article L. 122-4). Une telle représentation ou reproduction, par quelque procédé que ce soit, constituerait donc une contrefaçon sanctionnée par les articles L. 335-2 et suivants du Code de la propriété intellectuelle.

Sommaire

Chrysta PÉLISSIER, Sébastien IKSAL, Jean-Marie GILLIOT, Patricia MARZIN-JANVIER • Éditorial du numéro spécial « EIAH 2023 » 7

ARTICLES DE RECHERCHE

Mathieu MURATET, Sébastien JOLIVET • GTC un formalisme permettant l'analyse automatique des compétences de la pensée informatique dans un jeu sérieux : application au jeu SPY 10

Bérénice LEMOINE, Pierre LAFORCADE, Sébastien GEORGE • Framework de conception et d'implémentation de générateurs d'activités de jeu d'entraînement aux connaissances déclaratives 45

Madjid SADALLAH, Jean-Marie GILLIOT • Modéliser et outiller la conception participative et générative de tableaux de bord d'apprentissage soutenant la création de sens 82

Christine MICHEL, Laëtitia PIERROT • Vers la conception de moyens et méthodes fondés sur les modèles pour caractériser et diagnostiquer la maturité numérique des enseignants 116

Mélina VERGER, Francois BOUCHET, Sébastien LALLÉ, Vanda LUENGO • Discriminations intersectionnelles : approfondir l'évaluation de l'équité algorithmique. L'exemple de la prédiction de la réussite académique avec des données issues de cours en ligne 137

Anaëlle BADIER, Marie LEFEVRE, Mathieu LEFORT, Nathalie GUIN • *Adaptive Learning* en contexte parascolaire : comprendre les usages et effets via l'analyse des traces d'un déploiement industriel 163

Comités 199

Édito numéro spécial « EIAH 2023 »

Chrysta PÉLISSIER¹ ; Sébastien IKSAL² ; Jean-Marie GILLIOT³ ; Patricia MARZIN-JANVIER⁴

¹Université de Montpellier – LHUMAIN – Université Paul Valéry, Montpellier 3

²Le Mans Université – LIUM

³IMT Atlantique – Lab-STICC

⁴Université de Bretagne Occidentale – CREAD

1. INTRODUCTION : CONFERENCE EIAH 2023

Ce numéro spécial présente une sélection de six articles issus de la 11^e édition de la conférence EIAH (Environnements Informatiques pour l'Apprentissage Humain) qui s'est déroulée à Brest (France) du 12 au 16 juin 2023.

Organisée tous les deux ans, cette conférence francophone, parrainée par l'Association des Technologies de l'Information pour l'Éducation et la Formation (ATIEF), a pour objectif la diffusion de travaux de recherche et la formation des chercheurs effectuant des études scientifiques en lien avec les EIAH. Cette communauté pluridisciplinaire communique sur des recherches dont la finalité est de susciter et d'accompagner les activités d'enseignement et d'apprentissage intégrant l'informatique. L'enjeu est d'aider à la conception, au développement, à la coordination, et à l'évaluation des situations d'apprentissage, qu'elles soient en présence, à distance, hybrides, comodales, en contextes scolaire, universitaire ou professionnel (formation initiale ou continue). L'objectif est de fournir divers moyens de comprendre, d'explicitier et de faire évoluer les pratiques sociales, organisationnelles et professionnelles des différents acteurs impliqués dans ces démarches : enseignant/formateur/ingénieur pédagogique, élève/apprenant/alternant, institution/école/université.

2. PRESENTATION DES ARTICLES DU NUMERO

Les textes proposés sont des versions étendues d'articles publiés dans les actes de la conférence EIAH 2023. Leurs auteurs ont été sollicités par les co-présidents du comité de programme (Sébastien Iksal et Chrysta Pélissier) en collaboration avec la rédactrice en chef de la revue STICEF (Élise Lavoué), en fonction des notes attribuées aux articles par les deux relecteurs du comité de programme de la conférence. Sur neuf auteurs ou co-auteurs sollicités, six d'entre eux ont accepté de soumettre à la revue une version étendue de leur article.

Chaque soumission a suivi le processus de sélection imposé par la revue, à savoir une relecture par deux évaluateurs. L'un des deux relecteurs a été choisi parmi les membres du comité de lecture de la revue STICEF, l'autre parmi les membres du comité de programme de la conférence EIAH 2023. Afin de couvrir différentes disciplines, un des relecteurs était issu des Sciences Humaines et Sociales (SHS) et l'autre de l'Informatique. Suite aux différentes remarques des relecteurs et d'un méta-relecteur (l'un des deux co-présidents du comité de programme, Chrysta Pélissier et Sébastien Iksal ; ou l'un des deux co-présidents du

comité d'organisation, Jean-Marie Gilliot et Patricia Marzin-Janvier), tous les articles ont été retenus pour une publication dans ce numéro spécial après révisions.

Deux grandes problématiques sont abordées dans ces six articles, l'une en lien avec le processus de conception d'un EIAH, la deuxième avec son évaluation. Par ailleurs, deux domaines disciplinaires scientifiques sont représentés : les sciences de l'information et de la communication (un article) et l'informatique (cinq articles).

2.1 CONCEPTION D'UN COURS OU D'UN EIAH

Quatre contributions interrogent le processus de conception. La première propose de décrire les compétences de la pensée informatique dans un jeu sérieux. La seconde vise à élaborer un *framework* de conception et d'implémentation de générateurs d'activités, également dans le domaine des jeux sérieux. La troisième présente une démarche de modélisation et d'outillage liée à la conception participative et générative de tableaux de bord. Enfin, la dernière discute les modèles pour la caractérisation et le diagnostic de la maturité numérique des enseignants.

Mathieu Muratet et Sébastien Jolivet proposent une méthode de description des compétences de la pensée informatique dans les jeux sérieux. La problématique porte sur l'élaboration d'un formalisme visant à décrire les compétences travaillées par combinaison des fonctionnalités d'un jeu. Ce formalisme se donne pour objectif d'aider les enseignants de l'école primaire à aborder ces compétences trop souvent mal maîtrisées. Les résultats d'intégration du formalisme sont encourageants. Il apparaît très utile pour caractériser la plupart des compétences liées à la pensée informatique. L'expérimentation réalisée a permis d'identifier des détournements possibles de certaines fonctionnalités du jeu SPY à des fins pédagogiques.

Bérénice Lemoine, Pierre Laforcade et Sébastien George nous proposent une aide à la conception de générateurs d'activités adaptées, dans le domaine des jeux sérieux également. La problématique porte sur l'élaboration d'un *framework* de conception de générateurs d'activités, orienté modèles et méta-modèles, extensible à différents domaines didactiques. Les résultats soulignent la diversité des activités d'entraînement adaptées produites à l'aide de ce *framework*, notamment sous forme de niveaux, ainsi que son apport en matière d'évaluation du point de vue ingénierie.

Madjid Sadallah et Jean-Marie Gilliot se proposent de modéliser et d'outiller la conception participative et générative de tableaux de bord d'apprentissage soutenant la création de sens pour le concepteur du tableau de bord. La recherche interroge un cadre de conception participatif détaillant les dimensions d'un espace dédié à la conception des tableaux de bord d'apprentissage, mettant particulièrement l'accent sur les explications des concepteurs pour étayer la prise de décision de leurs utilisateurs. Les résultats démontrent que ces outils offrent un soutien significatif tant pour la participation des utilisateurs finaux (enseignants, apprenants, décideurs) que pour l'expérience des développeurs, soulignant ainsi leur pertinence et leur efficacité.

Christine Michel et Laëtitia Pierrot exposent les modèles visant la caractérisation et le diagnostic de la maturité numérique des enseignants. La revue de littérature des modèles de maturité numérique des enseignants réalisée offre une synthèse des dimensions constitutives de chacun des modèles. Elle précise aussi la manière dont ces dimensions contribuent à déterminer les niveaux de maturité numérique des enseignants. Du point de vue des résultats, la synthèse met en lumière une diversité des dimensions retenues, ainsi qu'une disparité de ces mêmes dimensions dans chacun des modèles. La dernière partie de l'article propose un modèle unifié de caractérisation et de diagnostic de la maturité numérique des enseignants.

2.2 L'ÉVALUATION DES EIAH

Deux articles abordent la thématique de l'évaluation des EIAH. Le premier propose une manière d'évaluer l'équité algorithmique vis à vis d'attributs sensibles et le second présente une analyse des usages et des effets d'une application mobile.

Mélina Verger, François Bouchet, Sébastien Lallé et Vanda Luengo se proposent d'évaluer l'équité algorithmique des modèles prédictifs utilisés en éducation en vue de prédire la réussite des étudiants à des cours en ligne. Prenant appui sur une théorie de l'intersectionnalité, les auteurs défendent l'idée que les influences réciproques de plusieurs attributs sensibles (genre, âge, handicap, niveau de pauvreté, ...) produisent des discriminations uniques et différentes pour certains sous-groupes d'individus. Les résultats mettent en lumière des discriminations algorithmiques (envers les hommes plus pauvres avec un handicap et plus âgés) qui n'étaient pas identifiables à partir des analyses classiques. Ils permettent aussi de déterminer l'influence de chaque attribut sur les discriminations grâce à leurs interactions avec les autres.

Anaëlle Badier, Marie Lefèvre, Mathieu Lefort et Nathalie Guin présentent une analyse de traces visant à identifier les usages et les effets d'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. Dans cette recherche, 8 000 apprenants ont été suivis sur 4 mois. Ils avaient des objectifs et des modalités de travail très variés, générés à partir d'un moteur de recommandations de ressources pédagogiques. L'analyse de traces d'utilisation de cette application souligne que les recommandations répondent bien à des objectifs et des usages différents et que leur suivi a une influence positive sur l'expérience d'apprentissage. Elle met également en évidence des leviers d'amélioration pour proposer un mécanisme de recommandations par conception itérative.

3. CONCLUSION ET PERSPECTIVES

Pour ce numéro spécial EIAH 2023, certains articles posent et interrogent des méthodes : de description des compétences qui prennent la forme d'un formalisme, de génération d'activités adaptées (*framework*) et de méthodes de génération de tableaux de bord.

D'autres articles abordent la notion de dimensions (constitutives de chacun des modèles, d'un espace dédié à la conception des tableaux de bord) et d'attributs. Ces dimensions et ces attributs amènent leurs auteurs à proposer un modèle unifié, une diversité des activités d'entraînement adaptées produites par l'EIAH, un soutien significatif tant pour la participation des utilisateurs finaux que pour l'expérience des développeurs.

La mise en œuvre de leur recherche amène les auteurs à observer des détournements possibles de l'EIAH, des apports en matière d'évaluation du point de vue ingénierie, une diversité de dimensions dans les modèles théoriques, des discriminations algorithmiques ainsi que des recommandations qui répondent bien à des objectifs et des usages pédagogiques différents.

Enfin, certains articles entrevoient par leurs études des leviers d'amélioration pour proposer un mécanisme de recommandation par conception itérative, ou visant à déterminer l'influence de chaque attribut sur les discriminations identifiées.

GTC un formalisme permettant l'analyse automatique des compétences de la pensée informatique dans un jeu sérieux : application au jeu SPY

GTC : A Formalism for the Automatic Analysis of Computational Thinking Skills in a Serious Game : Application to the SPY Game

Mathieu MURATET^{1,2}; Sébastien JOLIVET³

¹ Sorbonne Université, CNRS, LIP6, F-75005 Paris, France

² INS HEA, 92150 Suresnes, France

³ IUFE & TECFA, Université de Genève, Suisse

Résumé. La pensée informatique est un ensemble de compétences souvent mal maîtrisé par les enseignants de l'école primaire. Pour les aider à aborder ces compétences avec leurs élèves, nous développons le jeu SPY. Dans cet article, nous menons une analyse du jeu SPY et proposons le formalisme GTC (Game To Competency) pour décrire les compétences travaillées, à partir des fonctionnalités du jeu. GTC a été appliqué pour caractériser 21 des 26 compétences d'un référentiel (PIAF). Nous montrons comment ces caractérisations peuvent être exploitées pour construire des indicateurs sur les compétences mobilisées dans chaque mission. Enfin, nous présentons une expérimentation visant à évaluer l'utilité perçue des propositions faites. Elle a mobilisé cinq enseignants d'informatique, en formation à l'Université de Genève, dans le cadre d'un atelier d'échanges dirigés. Les résultats obtenus, très positifs, fournissent un premier niveau de validation des propositions faites et dégagent de nouvelles pistes de recherches.

Mots-clés : Jeu sérieux, pensée informatique, modélisation, compétence, analyse automatique, utilité

Abstract. *Computational thinking is a skillset inadequately mastered by elementary school teachers. To aid teachers in addressing these competencies with their students, we have developed the SPY game. In this paper, we conduct an analysis of SPY and we propose a formalism to describe the competencies targeted by the game based on its mechanics. The proposed formalism has been applied to describe 21 out of the 26 competencies within PIAF conceptual framework. We show how these descriptions allow for the construction of indicators elucidating the competencies involved. Finally, we present an experiment designed to evaluate the usefulness of the proposals made. Five teachers undergoing training at the University of Geneva participated in the experiment within the framework of a supervised workshop. The highly positive outcomes attained offer an initial validation of the proposed methods and suggest new research directions.*

Keywords: *Serious game, computational thinking, modeling, competency, automatic analysis, usefulness*

1. INTRODUCTION ET POSITIONNEMENT

Depuis maintenant quelques années, l'informatique est devenue une discipline intégrée aux programmes scolaires (Baron *et al.*, 2014), englobant un ensemble de compétences désigné sous le terme de « pensée informatique ». Selon Wing (2006), la pensée informatique mobilise cinq capacités cognitives : (1) la pensée algorithmique, (2) l'abstraction, (3) l'évaluation, (4) la décomposition, et (5) la généralisation. Sur cette base, Parmentier *et al.* (2020) proposent un découpage détaillé des compétences de la pensée informatique et algorithmique pour l'enseignement fondamental (le référentiel PIAF). La question de l'appropriation de ces compétences par les enseignants de l'école primaire reste complexe à résoudre. En effet, enseigner cette nouvelle discipline requiert un investissement particulièrement important de leur part (Kradolfer *et al.*, 2014), notamment en raison d'un manque de formation (initiale et continue). Ainsi, par manque de maîtrise de ces compétences, les enseignants éprouvent des difficultés pour élaborer leurs propres scénarios pédagogiques ou évaluer la pertinence des outils visant à développer ces compétences chez leurs élèves (Poirson, 2020).

Sigayret *et al.* (2021) identifient trois principaux dispositifs pour initier les élèves à la pensée informatique : les activités débranchées n'utilisant aucun outil numérique, l'approche tangible de la robotique éducative et l'utilisation de logiciels visuels de programmation.

L'informatique débranchée offre une porte d'entrée riche dans la science informatique, n'exigeant pas de matériels coûteux (Alayrangues *et al.*, 2017). Cependant, elle demande à l'enseignant une expertise approfondie sur le savoir abordé afin d'être en mesure de guider les élèves dans la construction de leurs connaissances. Ces mêmes auteurs (Alayrangues *et al.*, 2017) soulignent que la mise en place de telles activités nécessite des connaissances en informatique ainsi que des compétences en pédagogie. Ce « super prof », maîtrisant ce savoir informatique, reste donc une exception ce qui rend difficile la démocratisation de ce type de scénario pédagogique à l'école primaire.

En ce qui concerne l'approche tangible de la robotique pédagogique, les objets tangibles offrent un artefact fertile sur le plan cognitif pour développer des compétences mathématiques et des formes de pensée algorithmique (Komis et Misirli, 2011). Ces robots programmables sont utilisés dans des scénarios pédagogiques, exigeant de l'enseignant une maîtrise fine des savoirs en jeu pour soutenir les travaux des élèves. Spach (2019) souligne à ce propos les freins au développement de cette approche tangible : « *Le défaut de maîtrise conceptuelle des enseignants est sans doute à l'origine du manque de référencement et du déficit d'institutionnalisation des notions et des concepts abordés dans les situations pédagogiques* ».

Dans ces deux premières approches, l'expertise de l'enseignant sur les savoirs informatiques est donc fortement sollicitée mais fait souvent défaut comme nous avons pu l'évoquer précédemment (Alayrangues *et al.*, 2017 ; Spach, 2019). La troisième approche, qui exploite l'utilisation de logiciels visuels de programmation, peut être un élément de réponse à cette problématique car ces environnements peuvent intégrer des systèmes intelligents pour accompagner les élèves et les enseignants. Pour contribuer à répondre à cette problématique, nous proposons un jeu sérieux nommé SPY qui permet aux élèves de travailler les compétences de la pensée informatique et fournit aux enseignants des moyens de comprendre ce qui est travaillé par leurs élèves. Il existe plusieurs dizaines de jeux sérieux d'apprentissage sur le thème de l'informatique (Lindberg *et al.*, 2019 ; Miljanovic et Bradbury, 2018). Les activités qu'ils proposent sont généralement structurés autour des notions, mais les compétences mises en oeuvre au sein d'une activité sont rarement identifiées. Par ailleurs, ces jeux fournissent peu d'outils à destination des enseignants pour les aider à adapter les jeux à leur contexte (Saddoug *et al.*, 2022). Ce sont ces limitations qui ont motivé la conception du jeu sérieux SPY.

La problématique générale est la suivante : comment aider les enseignants peu familiers avec la pensée informatique à repérer dans des situations de jeu (qu'ils pourraient eux-mêmes créer) les compétences mobilisées ? Dans cet article nous présentons une contribution générale afin de lier des compétences aux fonctionnalités ludiques d'un jeu sérieux. Nous mettons en oeuvre cette contribution dans le jeu sérieux SPY.

Nous fixons dans un premier temps le cadre théorique ainsi que les questions de recherche dans la section 2. Dans la section 3, nous présentons le jeu SPY utilisé dans le cadre de cette étude et détaillons ses variables didactiques. La section 4 est consacrée à la présentation du formalisme GTC (Game To Competency) permettant de décrire les compétences de la pensée informatique à partir des fonctionnalités de jeu, et illustre l'application de ce formalisme pour décrire quelques compétences de la pensée informatique telles que mises en oeuvre dans SPY. Dans la section 5, nous montrons les outils qui ont pu être construits sur la base du formalisme GTC pour offrir à l'enseignant de nouveaux outils centrés sur les compétences de la pensée informatique. Enfin, dans la section 6, nous présentons une expérimentation visant à évaluer l'utilité perçue des outils proposées, dont les résultats sont exposés dans la section 7.

2. ANCRAGE THÉORIQUE ET QUESTIONS DE RECHERCHE

SPY est un jeu conçu pour être utilisé en classe sur la prescription d'un enseignant. Sa pertinence comme élément de réponse à notre problématique (outiller les enseignants, y compris non-experts du domaine, pour enseigner la pensée informatique) nécessite de l'examiner sous différents angles. Pour aborder cet examen nous différencions trois points de vues : celui des *concepteurs* du jeu et ceux des deux catégories d'utilisateurs finaux que sont les *enseignants* et les *apprenants*. Différentes approches théoriques permettent de décrire ou d'éclairer les choix et les effets selon ces trois points de vue. Dans cet article, en l'absence d'expérimentations en conditions écologiques, nous nous concentrons sur les deux premiers points de vue à savoir les *concepteurs* et les *enseignants*.

Tout d'abord, s'agissant de la conception d'un EIAH, en se focalisant sur la question du savoir dans l'environnement, Balacheff introduit en 1994 le concept de *transposition informatique* de la manière suivante : « *aux contraintes de la transposition didactique s'ajoutent, ou plutôt se combinent, celles de modélisation et d'implémentation informatiques : contraintes de la modélisation computable, contraintes logicielles et matérielles des supports informatiques de réalisation. Ce que l'on place habituellement sous le terme d'informatisation ne constitue pas une simple translittération, les environnements informatiques d'apprentissage résultent d'une construction qui est le lieu de transformations nouvelles des objets d'enseignement* » (Balacheff, 1994, p. 364).

De plus, nous travaillons sur un type d'EIAH particulier incluant une forte dimension ludique : un jeu sérieux. Dans notre situation les apprentissages sont placés au cœur de la jouabilité. Marne *et al.* (2012) parlent alors de jeu sérieux d'apprentissage à métaphore intrinsèque ce qui implique que les mécanismes de gameplay soient en cohérence avec les contenus à travailler. Ainsi, il existe une relation forte entre les fonctionnalités de jeu et le savoir à enseigner, elles ont donc, pour la plupart, une fonction didactique.

Dans la théorie de la médiation sémiotique, Mariotti et Maracci (2012) développent l'idée du potentiel sémiotique d'un artefact qui 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* » C'est précisément ce second lien qui est inféré lorsque

l'on fait l'hypothèse que l'usage d'une fonctionnalité du jeu (considérée comme artefact) peut être interprété comme relevant de la mobilisation d'un savoir de la pensée informatique.

Ainsi, lors de la conception d'un tel EIAH, relativement au savoir travaillé, se combinent des enjeux et processus de transpositions didactique, de transposition informatique et une forme de transposition ludique. Bonnat *et al.* (2023) proposent une étude de ce processus qu'ils nomment *ludicisation*.

Par ailleurs, les concepteurs sont amenés à réaliser différents choix, guidés ou contraints par : les objectifs d'apprentissage fixés à l'EIAH, leurs représentations sur l'apprentissage, les possibilités et limites technologiques, etc.

Bruillard (1997) note justement la complexité de conception d'EIAH et souligne la « *difficulté réelle d'incorporer les théories de l'apprentissage et de l'enseignement dans les systèmes informatiques à finalité éducative* ». Finalement, dans la très grande majorité des cas, même pour leur part identifiée et consciente, les choix et contraintes qui dirigent la conception d'un EIAH, ainsi que leurs conséquences sur l'apprentissage lors de leur utilisation, ne sont pas accessibles aux utilisateurs finaux, tels que les enseignants et/ou les apprenants dans notre cas.

Du point de vue de l'utilisateur final enseignant, la connaissance des choix et contraintes qui ont orienté la conception de l'EIAH n'est pas fondamentale. Par contre, il est pertinent qu'il puisse connaître les effets, de ces choix et des éléments qu'il peut manipuler, sur l'apprentissage.

L'un des enjeux est donc de décrire les effets et potentialités didactiques de l'EIAH et de rendre accessibles ces informations pour l'enseignant. Pour ce faire nous considérons tout d'abord que le jeu s'intègre dans une situation didactique (Brousseau, 1990) et que l'enseignant, en utilisant les potentialités du jeu, va pouvoir définir une situation, se voulant *a-didactique* (Brousseau, 1986), pouvant être traitée sans intervention de l'enseignant par l'élève et nécessitant la mobilisation des savoirs visés pour atteindre l'objectif du jeu. Il est donc nécessaire d'identifier dans le jeu les fonctionnalités pouvant être considérées comme des variables didactiques, c'est-à-dire : « *un élément de la situation sur lequel le maître peut agir ; qui provoque des changements qualitatifs dans les procédures de résolution des élèves ; qui permet d'expliquer les résultats de l'enseignement et d'agir sur eux ; et qui provoque une modification dans l'apprentissage* » (Margolinas, 1992, p. 129).

Ces différents éléments vont être exploités par l'enseignant lors du processus de genèse instrumentale (Rabardel, 1995) qui va faire passer l'EIAH du statut d'artefact au statut d'instrument pour son enseignement (Rabardel, 1999). Pour cela, l'enseignant va, en fonction des apprentissages visés, activer tout ou partie des fonctionnalités de l'artefact. Puis, l'élève-joueur va à son tour engager une genèse instrumentale en interagissant avec l'instance du jeu sérieux défini par l'enseignant, et développer des schèmes d'utilisation l'amenant à construire son propre instrument. Les schèmes pouvant être développés par l'élève dépendent des fonctionnalités, prédéfinies ou activées par l'enseignant dans l'artefact et de leur combinaison. Par exemple, dans le contexte d'un jeu de type « déplacement d'un robot dans un labyrinthe », limiter le nombre d'actions de déplacements et mettre à disposition une structure de contrôle permettant de réaliser des boucles bornées, vont permettre aux élèves de construire des schèmes liés à l'identification d'un schéma à répéter et du nombre d'itérations requis, ainsi qu'à l'agencement des différents composants du programme.

La capacité des enseignants à instrumentaliser l'EIAH pour qu'il instrumente au mieux leur enseignement dépend de divers facteurs parmi lesquels sa formation et sa maîtrise du savoir à enseigner, sur lesquelles nous ne revenons pas dans cet article. Mais aussi de la compréhension qu'ils peuvent avoir des effets, en termes de potentialités d'apprentissages, des choix de développements et des mécaniques de jeu implémentées. Dans cet article, nous

souhaitons donc expliciter et exemplifier, pour l'utilisateur final *enseignant*, ces effets et potentialités dans SPY. Ceci se concrétise par deux questions de recherche :

1. Comment lier les compétences d'un référentiel avec les fonctionnalités ludiques d'un jeu sérieux ?
2. Comment exploiter ce lien pour analyser automatiquement les missions du jeu sérieux SPY et en extraire les compétences de la pensée informatique ?

3. ANALYSE DU JEU SPY

SPY¹ (voir Figure 1) est un jeu sérieux d'apprentissage sur le thème de la pensée informatique. Il est conçu pour un public d'élèves de cycle 3 (9, 10, et 11 ans). C'est un projet *open source*² et *open data*³ développé au sein de Sorbonne Université. Le principe du jeu est de programmer un ou plusieurs robots à l'aide d'actions afin de les faire sortir d'un labyrinthe, c'est-à-dire de les déplacer du point de départ au point d'arrivée du labyrinthe. Ces actions, disponibles dans un inventaire, sont représentées sous forme de blocs que le joueur doit agencer en séquences dans la/les zone(s) d'édition. Il s'agit donc, *in fine*, de réaliser un programme dans un langage par blocs.

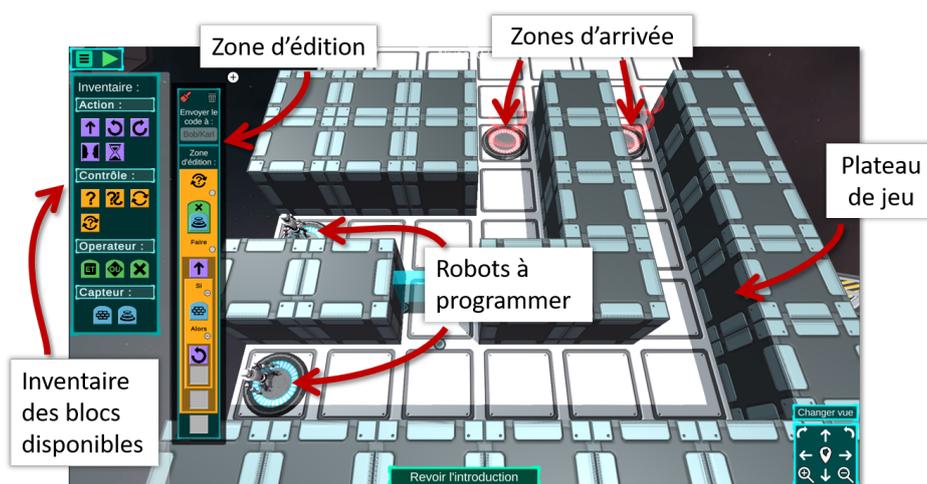


Figure 1 : Interface joueur commentée d'une mission de SPY

SPY est conçu comme une ressource à l'interface de la robotique pédagogique et des environnements visuels de programmation. Du point de vue du joueur, SPY reprend les codes de la robotique pédagogique où des actions simples de déplacement, représentées sous forme de blocs, permettent de programmer un ou plusieurs robots. Ces blocs sont complétés par des blocs plus complexes (incluant des structures de contrôle, des opérateurs logiques et des capteurs) utiles pour résoudre certaines situations du jeu. Cela permet aux élèves une première rencontre avec ces nouveaux concepts avant de les manipuler dans des environnements de programmation plus ouverts comme Scratch, puis dans le cadre de la programmation textuelle.

Du point de vue de l'enseignant, SPY fournit deux éditeurs pour leur permettre d'adapter la ressource à l'usage souhaité :

1. SPY : <https://spy.lip6.fr>
2. Code source du jeu SPY : <https://github.com/Mocahteam/SPY>
3. Données ouvertes du jeu SPY : <https://spy.lip6.fr/openTraces.html>

- *L'éditeur de mission* permet à l'enseignant de créer ses propres missions ou de modifier les missions déjà intégrées au jeu.
- *L'éditeur de scénario* permet à l'enseignant d'agencer et scénariser des missions afin de construire sa propre séquence pédagogique.

La suite de cette section est organisée en deux parties. Tout d'abord, dans la section 3.1, nous présentons de manière détaillée les différents aspects de SPY qui lient les caractéristiques de l'artefact, notamment les ressorts ludiques, avec les notions relatives à la pensée informatique. Puis, dans la section 3.2, nous analysons certaines caractéristiques du jeu en termes de variables didactiques et/ou des éléments jouant sur la complexité des situations.

3.1. CONSTITUANTS DU JEU SPY ET LEURS FONCTIONS

3.1.1. Blocs de programmation (instructions et expressions)

Les blocs de programmation sont les éléments fondamentaux de SPY, car ils constituent les briques de base permettant d'écrire les programmes exécutés par les robots. Ils sont répartis en quatre catégories :

- **Les blocs d'action** : ils permettent de définir les actions pouvant être réalisées par les agents : « Avancer », « Pivoter à gauche », « Pivoter à droite », « Attendre », « Activer un terminal » et « Faire demi-tour ». À noter que toutes ces actions sont atomiques à l'exception de l'action « Faire demi-tour » qui peut être décomposée en deux actions « Pivoter à droite » ou deux actions « Pivoter à gauche ».
- **Les blocs de contrôle** : ils permettent de contrôler les blocs d'action à exécuter : « Si Alors », « Si Alors Sinon », « Répéter n fois », « Tant que » et « Répéter indéfiniment ».
- **Les capteurs** : ils 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 interdite ou une porte se trouve en face de lui ; et si un terminal ou une sortie se trouve sur sa position.
- **Les opérateurs** : ils 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 ».

Les blocs de programmation sont liés aux notions d'instruction et d'expression. Les choix réalisés en termes de transposition de ces notions dans SPY portent notamment sur les formes associées aux différents blocs. Ainsi, dans le jeu, les instructions (blocs d'action et de contrôle) sont représentées par des formes à l'entête carrée alors que les expressions (blocs capteurs et opérateurs) sont représentées par des formes à l'entête arrondie. Cette représentation visuelle donne implicitement des informations sur la grammaire du langage et donc sur les blocs combinables ou non. La figure 2 présente la grammaire du langage de SPY, nous observons deux types de réceptacles de couleur grise, l'un avec une forme carrée et l'autre avec une entête arrondie. Les instructions de couleur violette (comme l'action « Avancer ») et les structures de contrôle de couleur orange (comme les blocs « Répéter n fois » et « Tant que ») sont de forme carrée et peuvent donc être glissés dans n'importe quel réceptacle carré. Les constituants d'une expression tels que les capteurs de couleur bleu ciel (comme le bloc « Sortie détectée ») et les opérateurs de couleur verte (comme le bloc « Non ») ont une entête de forme arrondie et peuvent donc être glissés dans n'importe quel

réceptacle aux bords supérieurs arrondis.

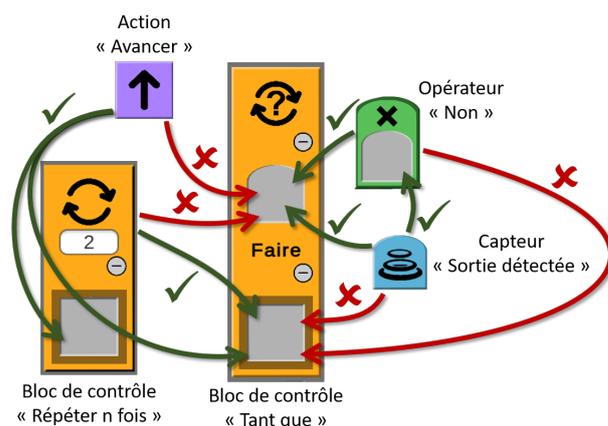


Figure 2 : Illustration de la grammaire des blocs de SPY

Ces différents blocs sont organisés, par glisser/déposer, dans une zone d'édition que nous présentons maintenant.

3.1.2. Zone d'édition

Une zone d'édition accueille les blocs de programmation permettant de construire, pour un robot, les solutions aux problèmes posés. Lorsque le problème implique plusieurs robots, chaque robot a sa propre zone d'édition associée. Une zone d'édition doit contenir au moins une instruction pour pouvoir être exécutée.

À chaque zone d'édition est associé le nom d'un robot, celui qui exécute la séquence d'instructions définie. Ceci fait référence à la notion de nommage et au fait qu'un objet informatique peut être référencé par son nom. Par exemple, dans la figure 1, si le nom du robot indiqué dans le champ « Envoyer le code à : » correspond au nom d'un des robots présents dans la scène, alors le code lui est envoyé pour qu'il l'exécute. Le résultat de cette exécution est visible sur le plateau de jeu.

3.1.3. Plateau de jeu

Dans SPY, le plateau de jeu est défini par un plan pouvant contenir différents éléments de jeu tels que du sol, des murs, des robots programmables, des points d'arrivée, des sentinelles, des portes, etc. Le joueur doit donc programmer un ou plusieurs robots pour les aider à atteindre les points d'arrivée tout en gérant les obstacles.

Les sentinelles, par exemple, interdisent des zones du labyrinthe dans leur ligne de vue. Si un robot se trouve sur l'une de ces zones, il est détecté, ce qui met fin à la partie. Le joueur doit donc programmer ses robots pour qu'ils évitent ces zones interdites. Tout comme pour les robots, les sentinelles peuvent contenir un programme pré-construit par l'enseignant (que le joueur ne peut modifier) ce qui rend les zones interdites dynamiques. Le joueur doit donc sélectionner chaque sentinelle du jeu, vérifier si un programme lui est affectée et, le cas échéant, observer le programme, le comprendre, anticiper les mouvements de la sentinelle et programmer son robot en conséquence.

Les portes sont également des objets interactifs. Elles ont deux états (ouvert ou fermé) et sont contrôlées par des terminaux de contrôle. Les portes et les terminaux de contrôle sont donc un premier exemple de transposition informatique des notions d'état, d'événement et

d'objet modifiable. Ainsi, agir sur un terminal de contrôle a comme conséquence de modifier l'état d'une porte (l'ouvrir ou la fermer).

3.1.4. Outils de contrôle sur l'exécution des programmes

Lorsque le joueur souhaite tester sa solution, il clique sur un bouton pour lancer l'exécution. Chaque zone d'édition envoie son programme au robot ciblé s'il existe, et chaque agent (robot ou sentinelle) exécute ses actions en parallèle. Le joueur peut suivre l'exécution des programmes (voir Figure 3) en observant les agents bouger sur le plateau de jeu en fonction des actions en cours d'exécution (mises en évidence dans le panneau d'exécution des programmes). Le joueur peut mettre en pause l'exécution à tout moment et exécuter les programmes pas à pas. Plusieurs savoirs informatiques sont représentés par ces choix.

Tout d'abord, le choix de l'utilisation d'un panneau d'affichage, appelé « contexte d'exécution » (voir Figure 3), différent de la zone d'édition (voir Figure 1), matérialise la différence entre un programme en cours d'édition et ce même programme chargé en mémoire par les robots et exécuté. Dans l'exemple de la figure 3, un même programme a été envoyé à deux robots différents qui exécutent donc leur programme indépendamment l'un de l'autre. Un premier robot est en train d'exécuter l'action « Pivoter à gauche » car il était face à un mur, alors que le second exécute l'action « Avancer ».

Ensuite, la mise en évidence du résultat de l'évaluation des capteurs renvoie à la notion d'évaluation d'expressions booléennes. Ici, seuls les capteurs notifient le résultat de leur évaluation et non les opérateurs logiques. C'est donc au joueur de reconstruire l'évaluation globale de l'expression booléenne à partir des indices donnés sur chaque capteur.

Enfin, plus classique, la notion de contrôle de l'exécution d'un programme est introduite avec un panneau dédié permettant de mettre en pause l'exécution du programme, d'exécuter pas à pas, de reprendre l'exécution, de l'accélérer ou de l'interrompre.

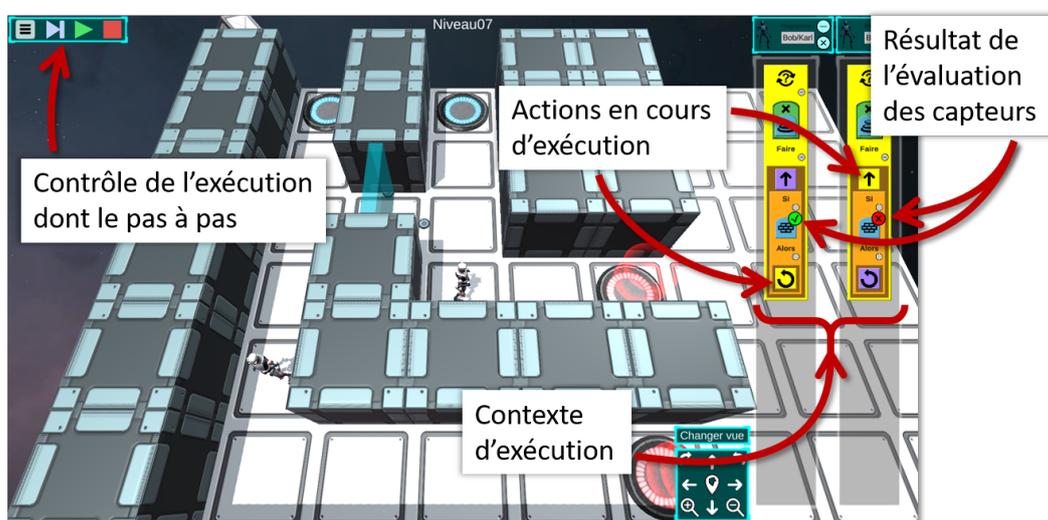


Figure 3 : Vue du jeu SPY en contexte d'exécution

3.1.5. Brouillard et texte d'introduction

Dans une mission 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 de la mission décrit dans les textes d'introduction. Le brouillard permet de modifier cette règle en limitant la vue d'un agent à son entourage proche (voir Figure 4). Dans ce cas, le texte d'introduction

présentant l'objectif de la mission joue un rôle fondamental, car il doit 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 ou sous la forme d'un algorithme, et le joueur doit le traduire à l'aide du langage formel du jeu. Cette fonctionnalité permet donc de travailler explicitement la conversion de l'information donnée dans un premier registre vers le registre du langage de programmation.

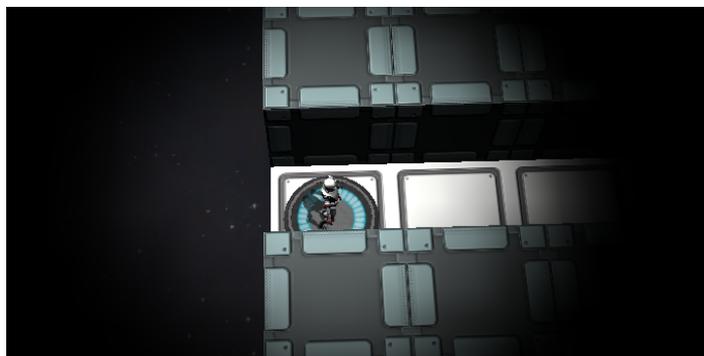


Figure 4 : Exemple de mission où le brouillard est activé.

3.2. VARIABLES DIDACTIQUES ET COMPLEXITÉ

Après avoir présenté les différents éléments constituant le jeu SPY et leur fonction par rapport aux savoirs travaillés nous reprenons certains de ces constituants en les examinant en tant que variables didactiques (Margolinas, 1992). Nous les considérons comme telles dans la mesure où, par le biais de l'éditeur de mission, il s'agit d'éléments pouvant être modifiés par l'enseignant et ayant un effet sur le processus de résolution que les élèves vont adopter pour résoudre les missions. Certains éléments peuvent également, sans nécessairement modifier les procédures de résolution efficaces ou disponibles, agir sur la complexité de l'exercice, par exemple en introduisant des choix à la charge de l'élève.

3.2.1. Blocs de programmation

Chaque bloc de programmation de l'inventaire peut être paramétré pour définir, dans une mission donnée, les blocs disponibles et en quelle quantité. Ainsi, l'accès à certains blocs (ou leur quantité) peut être un moyen pour forcer le joueur à utiliser certains blocs (par exemple, demander de faire avancer un robot de plusieurs cases en limitant le nombre de blocs « Avancer » à 1 et en donnant accès au bloc de contrôle « Répéter n fois »).

Ce paramétrage peut être utilisé pour réduire ou augmenter la complexité d'une mission. Par exemple, une mission d'introduction pourra ne contenir que les blocs utiles à sa résolution, évitant ainsi au joueur de devoir choisir parmi des blocs potentiellement inutiles.

3.2.2. Zone d'édition

Par défaut, chaque robot est associé à une zone d'édition. Il est cependant possible de rompre cette association afin de demander au joueur de la réaliser lui-même. Dans ce cas, le joueur doit indiquer dans la zone d'édition le nom de l'agent auquel elle est associée afin que le programme lui soit envoyé lors du lancement de l'exécution.

Une zone d'édition peut également être préconstruite. Le programme ainsi proposé au joueur peut être complet, partiel ou bogué. Le joueur devra le compléter ou le corriger en conséquence.

SPY offre également la possibilité de ne proposer au joueur qu'une seule zone d'édition pour plusieurs robots. Ceci permet de construire des missions où le joueur devra imaginer un unique programme permettant de contrôler plusieurs robots pouvant être dans des labyrinthes différents. Le joueur pourra ainsi se confronter à la réalisation de solutions plus génériques et par exemple, utiliser des boucles non bornées plutôt que des boucles bornées.

Enfin, le glisser/déposer est le mode d'interaction permettant d'ajouter des blocs de l'inventaire à une zone d'édition. Il est possible de désactiver cette fonctionnalité, dans ce cas la mission doit proposer un ou plusieurs programmes préconstruits que le joueur pourra exécuter sans pouvoir les modifier (voir Figure 5). Si cette variable didactique est activée, elle modifie profondément le type de tâche à réaliser par l'élève, en effet, il ne s'agira plus ici de construire un programme mais au contraire de lire et comprendre des programmes préconstruits.

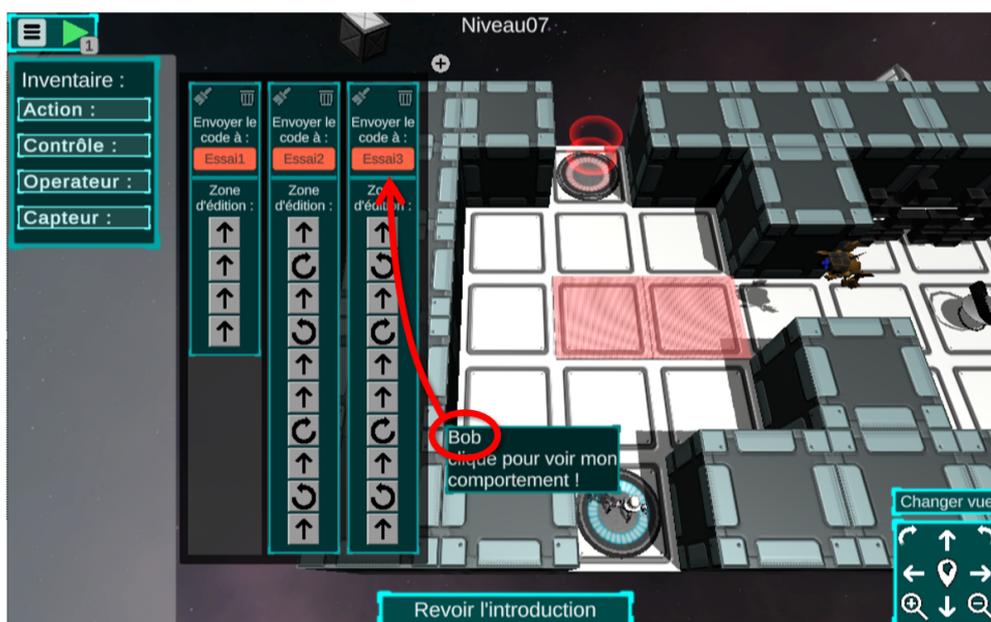


Figure 5 : Exemple de mission où le glisser/déposer est désactivé. Le joueur doit remplacer l'un des textes « Essai1 », « Essai2 » ou « Essai3 » avec le nom du robot (Bob). Dans cet exemple la bonne solution est « Essai3 ».

3.2.3. Plateau de jeu

Concernant le plateau du jeu, l'enseignant peut modifier sa taille ainsi que le contenu qui le compose. Il peut donc construire des labyrinthes plus ou moins complexes. Par exemple, si l'enseignant positionne l'arrivée à l'extérieur du champ de vue de la caméra, l'élève devra effectuer des actions de déplacement de la caméra afin de recueillir des informations sur la constitution d'une mission avant de commencer à réfléchir à la résolution du problème. Il pourra se poser des questions telles que : où se trouve l'arrivée ? Y a-t-il une sentinelle dans la mission ? etc. Il devra donc mobiliser des procédures relatives à la compétence consistant à comprendre les données du problème avant de tenter de le résoudre.

3.2.4. Contrôle de l'exécution des programmes

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 missions contenant des sentinelles en mouvement. SPY permet donc de résoudre une mission en plusieurs coups. Le joueur peut définir une première séquence de blocs, l'exécuter, observer la nouvelle situation, ajouter de nouveaux blocs, exécuter à nouveau, observer, etc. Il est donc possible de construire une solution étape par étape, ce qui permet aux élèves d'engager un processus de résolution par décomposition du problème.

Cependant, il peut être pertinent de limiter ce nombre d'exécutions pour amener progressivement le joueur à anticiper plusieurs actions à l'avance. Il est donc possible, pour chaque mission de SPY, de définir le nombre autorisé d'exécutions pour le résoudre.

3.3. SYNTHÈSE DES LIENS ENTRE FONCTIONNALITÉS DE JEU ET SAVOIRS

Tableau 1 : Synthèse des liens entre les fonctionnalités du jeu et les savoirs abordés par le jeu

		Fonctionnalités ludiques												
		Forme des blocs	Blocs « Se retourner » VS « Pivoter gauche/droite »	Bloc « Répéter n fois »	Bloc « Tant que »	Blocs « Si Alors » et « Si Alors Sinon »	Zone d'édition	Plateau de jeu	Portes/Terminaux	Contrôle de l'exécution	Brouillard	Sentinelles		
Savoirs	Grammaire	X												
	Instructions	X												
	Actions (non) atomiques		X											
	Répétition			X	X									
	Test logique				X	X								
	Expressions	X			X	X								
	Edition de programme						X							
	Lecture de code						X						X	
	Compréhension de code						X						X	
	Nommage						X							
	Comprendre les données du problème							X						
	Etat et évènements								X					
	Débogage									X				
	Traduction										X			
	Généricité						X							
Décomposition de problèmes									X					

L'analyse des différents constituants de SPY montre que chacun a une fonction didactique. Si certaines sont évidentes, comme les blocs de contrôle qui permettent de manipuler les notions de programmation associées, d'autres sont moins évidentes, comme l'association

de programmes aux sentinelles qui demande au joueur de comprendre une séquence d'actions et d'anticiper son exécution, ou l'utilisation d'une seule zone d'édition pour contrôler deux robots qui force le joueur à généraliser sa solution. Le tableau 1 présente une vue synthétique des liens entre les mécaniques de jeu et les savoirs liés.

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 (2006). 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 à de nouveaux problèmes (généralisation). Mais comment identifier, d'un point de vue micro, si telle ou telle compétence est en jeu dans une mission donnée ? Quelles influences les combinaisons de fonctionnalités ont-elles sur les compétences en jeu ? Répondre à ces questions est l'objet de la section suivante.

4. RÉTRO-INGÉNIERIE DES COMPÉTENCES À PARTIR DES FONCTIONNALITÉS DE JEU : LE FORMALISME GTC

Nous avons montré dans la section 3 comment les savoirs relatifs à la pensée informatique sont transposés dans le jeu SPY. Nous avons également évoqué quelques variables didactiques en relation avec les briques élémentaires du jeu. Dans cette section, nous proposons le formalisme GTC (Game To Competency) afin de décrire le potentiel sémiotique du jeu tel que défini par Mariotti et Maracci (2012) et ainsi répondre à notre première question de recherche : comment lier les compétences d'un référentiel avec les fonctionnalités ludiques d'un jeu sérieux ?

Le formalisme GTC que nous proposons pour répondre à cette question s'appuie sur le métalangage de description générique XML qui est ici utilisé pour décrire les missions d'un jeu. Si une mission contient une fonctionnalité de jeu donnée, la balise XML associée sera ajoutée au fichier XML décrivant la mission. Une mission est donc définie par l'agrégation d'un ensemble de balises XML, chacune décrivant l'état d'une fonctionnalité de jeu et de ses paramètres. Identifier une compétence dans une mission, consiste donc à repérer dans la structure XML de cette mission les fragments XML caractérisant cette compétence. Nous proposons donc de décrire chaque compétence de la manière suivante :

- le nom de la compétence ;
- une liste de filtres, chaque filtre correspondant au nom d'une balise XML et pouvant être complété par une liste de contraintes ;
- une règle qui définit les conditions globales que la compétence doit respecter. Cette règle est une expression booléenne qui combine les résultats des filtres.

Un filtre recherche, dans la structure XML décrivant une mission, toutes les occurrences d'une balise XML spécifique qui vérifient les contraintes définies. L'ensemble des contraintes possibles est présenté dans le Tableau 2.

En résumé, une compétence est décrite avec des filtres sur les balises XML et une règle globale qui orchestre ces filtres. Si cette règle est vraie, alors nous considérons que la compétence est mobilisée dans la mission analysée. Ces filtres s'appuient sur la structure du langage XML et sont indépendants des balises définies dans le jeu SPY. Notre proposition est donc générique pour tout jeu décrivant ses données à l'aide du langage de balisage XML. Trois exemples détaillés sont présentés dans les sections 4.1, 4.2 et 4.3.

L'identification d'une compétence repose donc sur l'évaluation d'un ensemble de filtres. Dans le contexte du jeu SPY, nous présentons dans le Tableau 3 les principales balises XML

Tableau 2 : Description des contraintes possibles sur un filtre.

$TAG \in$ ensemble des balises XML de SPY
 $ATTR \in$ ensemble des attributs valides pour TAG
 $OPP \in \{=, <, \leq, >, \geq, \neq\}$
 $VAL \in \mathbb{N}$

Contraintes applicables sur un filtre ciblant TAG	Description
\emptyset	Filtre les balises TAG .
$ATTR\ OPP\ VAL$	Filtre les balises TAG qui ont un attribut $ATTR$ égal/différent... à une valeur entière.
$ATTR\ isIncludedIn\ SET$	Filtre les balises TAG qui ont un attribut $ATTR$ dont la valeur est incluse dans un ensemble de valeurs (SET).
$ATTR\ sameValue\ TAG_2\ ATTR_2$	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 .
$hasChild$	Filtre les balises TAG qui contiennent au moins une balise fille.

que nous référençons dans cet article. Un modèle de mission complet et commenté est donné dans Muratet (2023a).

Tableau 3 : Description des principales balises structurant une mission SPY.

Tag	Description
<dragdropDisabled>	Si présente, désactive la fonctionnalité de glisser/déposer
<blockLimit block-Type="X" limit="Y">	Si présente, définit la quantité Y de blocs de type X disponible dans l'inventaire (si Y = -1 \Rightarrow bloc en quantité illimitée).
<script outputLine="X" editMode="Y" type="Z">	Définit une zone d'édition qui enverra son contenu sur le canal de communication X (voir balise <robot> et <guard>). La propriété editMode indique si le joueur peut changer ou non le canal de communication. La propriété type indique si cette zone d'édition contient un programme préconstruit optimal, non optimal, bogué ou indéfini.
<robot inputLine="X">	Définit un robot contrôlé par le joueur. Ce robot écoute le canal de communication X.
<guard inputLine="X">	Définit une sentinelle. Cette sentinelle écoute le canal de communication X.

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 (PIAF Project, 2021). 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 indique la sous-compétence Y de

la compétence X, 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 filtres construits pour caractériser les compétences du PIAF en fonction des fonctionnalités de jeu est accessible dans Muratet (2023b). Nous décrivons ici, en détail, l'application de GTC sur une sélection de trois compétences du PIAF et nous résumons les autres en Annexe 8.

4.1. 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 » (PIAF Project, 2021, p. 2).

Dans SPY, nous considérons que cette compétence est en jeu lorsque le joueur doit associer une zone d'édition à un robot. En effet, dans ce cas, il doit nommer son programme afin qu'il soit interprété par le bon robot.

Nous caractérisons cette compétence à l'aide d'un filtre sur la balise « script ». Nous complétons ce filtre avec une contrainte « **ATTR OP VAL** » qui est instanciée ainsi : « **editMode = 2** ». Ce filtre, noté $F1$, définit donc l'ensemble des balises « script » qui ont un attribut « editMode » égal à la valeur 2.

La règle complète pour cette compétence est alors $C1.1 := Card(F1) \geq 1$ qui signifie : « La compétence C1.1 est en jeu dans la mission si le cardinal de l'ensemble des balises sélectionnées par le filtre est supérieur à 0 », autrement dit « si la mission contient au moins une zone d'édition nommable ».

4.2. 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 » (PIAF Project, 2021, p. 5).

Dans SPY, cette compétence est travaillée lorsqu'une mission 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 une mission qu'au moins une zone d'édition est associée à une sentinelle et qu'elle contient au moins un bloc de programmation. Nous caractérisons cette compétence à l'aide d'un filtre sur la balise « script ». Nous notons ce filtre $F1$ et nous le complétons avec deux contraintes :

- $C1$ (de type « **ATTR sameValue TAG2 ATTR2** ») définit l'ensemble des zones d'éditions (« script ») qui communiquent sur un canal de communication (ATTR = « outputLine ») identique à celui d'une sentinelle (TAG2 = « guard » et ATTR2 = « inputLine »). $C1$ est donc instanciée ainsi : « **outputLine sameValue guard inputLine** » ;
- $C2$ (de type « hasChild ») définit l'ensemble des zones d'éditions (« script ») qui contiennent au moins un bloc de programmation (« hasChild »).

$F1$ peut donc être formellement défini comme l'intersection de $C1$ et $C2$, $F1 := C1 \cap C2$.

Ainsi vérifier dans une mission qu'au moins une zone d'édition est associée à une sentinelle et qu'elle contient au moins un bloc de programmation se résume avec la règle suivante : $C1.5 := Card(F1) \geq 1$.

4.3. 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 » (PIAF Project, 2021, p. 7).

Dans SPY, il s’agit d’une situation où il n’est proposé au joueur que des blocs en quantité limitée. Le joueur doit alors les combiner dans le bon ordre pour résoudre la mission. Pour décrire cette compétence, nous définissons les filtres suivants : (*F1*) le glisser/déposer est activé pour permettre au joueur de combiner les actions ; (*F2*) aucun bloc n’est disponible en quantité illimitée ; et (*F3*) il y a au moins une action disponible dans l’inventaire.

Le filtre *F1* porte sur la balise « dragdropDisabled ». Nous souhaitons vérifier que le glisser/déposer est actif et donc que la balise « dragdropDisabled » est absente dans la description de la mission. *F1* définit donc l’ensemble des balises « dragdropDisabled » et nous chercherons à vérifier dans la règle globale que cet ensemble est vide.

Le filtre *F2* porte sur la balise « blockLimit ». Nous complétons ce filtre avec une contrainte « **ATTR OP VAL** » qui est instanciée ainsi : « **limit = -1** ». *F2* définit donc l’ensemble des balises « blockLimit » qui ont un attribut « limit » égal à la valeur -1. Ce filtre définit l’ensemble des blocs disponibles en quantité illimitée, nous chercherons donc à vérifier dans la règle globale que cet ensemble est vide.

Le filtre *F3* porte lui aussi sur la balise « blockLimit ». Nous complétons ce filtre avec deux contraintes :

- *C1* (de type « **ATTR OP VAL** ») définit l’ensemble des blocs de programmation (« blockLimit ») dont le nombre d’exemplaires (ATTR = « limit ») est supérieur ou égal (OP = « \geq ») à 1 (VAL = 1). *C1* est donc instanciée ainsi : « **limit \geq 1** » ;
- *C2* (de type « **TAG ATTR isIncludedIn SET** ») définit l’ensemble des blocs de programmation (« blockLimit ») dont le type (ATTR = « blockType ») est inclus dans l’ensemble des blocs d’action (SET = {*Forward, TurnLeft, TurnRight, Wait, Activate, TurnBack*}). *C2* est donc instanciée ainsi : « **blockType isIncludedIn {Forward,TurnLeft,TurnRight,Wait,Activate,TurnBack}** ».

F3 peut donc être formellement défini comme l’intersection de *C1* et *C2*, $F3 := C1 \cap C2$.

Ainsi, vérifier dans une mission qu’aucun bloc d’action n’est donné en quantité illimitée se résume avec la règle suivante, $C2.1 := F1 = \emptyset \ \&\& \ F2 = \emptyset \ \&\& \ Card(F3) \geq 1$. À noter ici que nous cherchons bien à vérifier que les ensembles *F1* et *F2* sont vides à savoir que le glisser/déposer n’est pas désactivé et qu’il n’y a pas de blocs en quantité illimitée.

5. INTÉGRATION DU MODULE D’ANALYSE AUTOMATIQUE

Pour répondre à notre première question de recherche, nous avons proposé GTC, un formalisme générique pour décrire des compétences à l’aide de fonctionnalités ludiques de jeu. Nous avons détaillé la description de trois compétences du PIAF. Parmi les 26 compétences du PIAF, 21 peuvent être appliquées à SPY et sont résumées dans l’Annexe 8. En plus des compétences du PIAF, nous avons confronté GTC à un autre référentiel, nous avons ainsi décrit les 5 niveaux de la compétence « 3.4 Programmer » du domaine 3 « Création de contenus » du CRCN⁴ (Cadre de Référence des Compétences Numériques). Enfin, nous

4. CRCN, p. 13, <https://eduscol.education.fr/document/20389/download>, consulté le 08/11/2023

Tableau 4 : Analyse d'une même mission avec les différents référentiels

 <p>Textes de briefing : Bien... Karl vient de se téléporter dans le bâtiment juste devant les portes d'entrée. Il faut continuer à progresser, pour l'instant tout semble calme. La connexion avec le robot étant de bonne qualité, vous pouvez envoyer plusieurs ordres en même temps. Il suffit de les mettre les uns sous les autres dans la zone de programme.</p>	<p>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 C5.1 - Représenter des objets ou séquences d'actions au moyen d'une représentation formelle C6.1 - Vérifier si une séquence d'actions atteint un objectif donné</p>
	<p>Référentiel CRCN : Niveau 1</p>
	<p>Référentiel SPY : F1 - Résoudre un problème en plusieurs étapes F10 - Blocs limités F11 - Bloc action « Avancer »</p>

avons exploité GTC pour décrire l'ensemble des dimensions ludiques de SPY (dit référentiel SPY), ici une dimension ludique est souvent associée à une, et une seule, fonctionnalité ludique mais certaines dimensions nécessitent plus de finesse. Par exemple, décrire le principe d'ouverture et de fermeture de porte, ne se limite pas à la présence de portes dans le niveau (qui ne peuvent être que des éléments de décors si elles ne sont pas contrôlables). Dans ce cas, nous décrivons à l'aide de notre formalisme GTC, que le principe ludique de porte interactive est conditionné par la présence d'une porte connectée à un terminal et la présence d'au moins un bloc « Activer un terminal » dans l'inventaire du joueur. Toutes ces caractérisations sont accessibles dans Muratet (2023b).

Nous sommes ainsi en mesure d'analyser automatiquement chaque mission (ainsi que de nouvelles missions créées par l'enseignant) en fonction de différents référentiels. Le Tableau 4 illustre l'analyse d'une mission selon les trois référentiels. Cette visualisation fournit à l'enseignant des informations selon le référentiel de son choix. Dans cet exemple, le niveau 1 de la compétence 3.4 du CRCN est identifié. Cette compétence est définie comme « Lire et construire un algorithme qui comprend des instructions simples ». Le CRCN est moins précis que le PIAF mais offre une vue plus globale des compétences en jeu.

5.1. INTÉGRATION DE L'ANALYSEUR À SPY

Le jeu SPY est composé de 4 modules : un *launcher*, un *player* et deux éditeurs (voir Figure 6). Le *launcher* est le point d'entrée pour les élèves, il leur propose de choisir la mission à jouer à l'intérieur d'un scénario. Lorsque l'élève a sélectionné la mission du scénario qu'il souhaite jouer, le *launcher* lance le *player* qui exécute la mission sélectionnée.

Les enseignants, quant à eux, peuvent manipuler deux éditeurs pour créer ou modifier des scénarios ainsi que des missions de jeu. Afin de répondre à la seconde question de recherche (Comment exploiter ce lien [GTC] pour analyser automatiquement les missions du jeu sérieux SPY et en extraire les compétences de la pensée informatique?), nous avons intégré

notre analyseur de compétence à deux niveaux :

- dans le *launcher*, lors de la sélection d'un scénario, il informe l'élève et indirectement l'enseignant sur les compétences mobilisées dans ce scénario ;
- dans l'éditeur de scénario, selon une double modalité, il filtre les missions par compétence et analyse les compétences d'une mission sélectionnée.

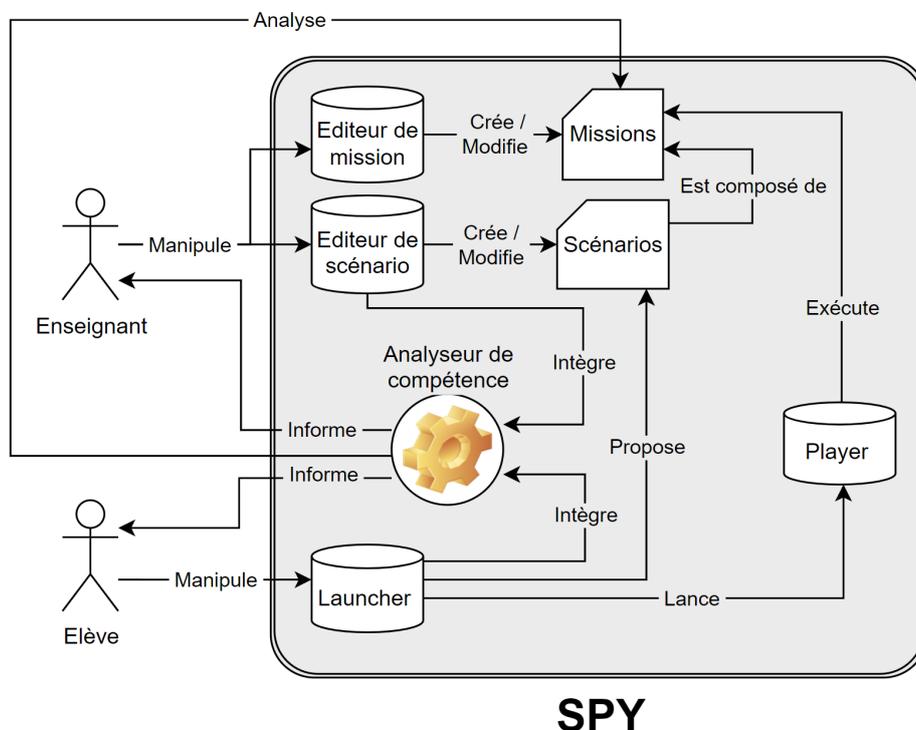


Figure 6 : Vue schématique de l'architecture logicielle de SPY

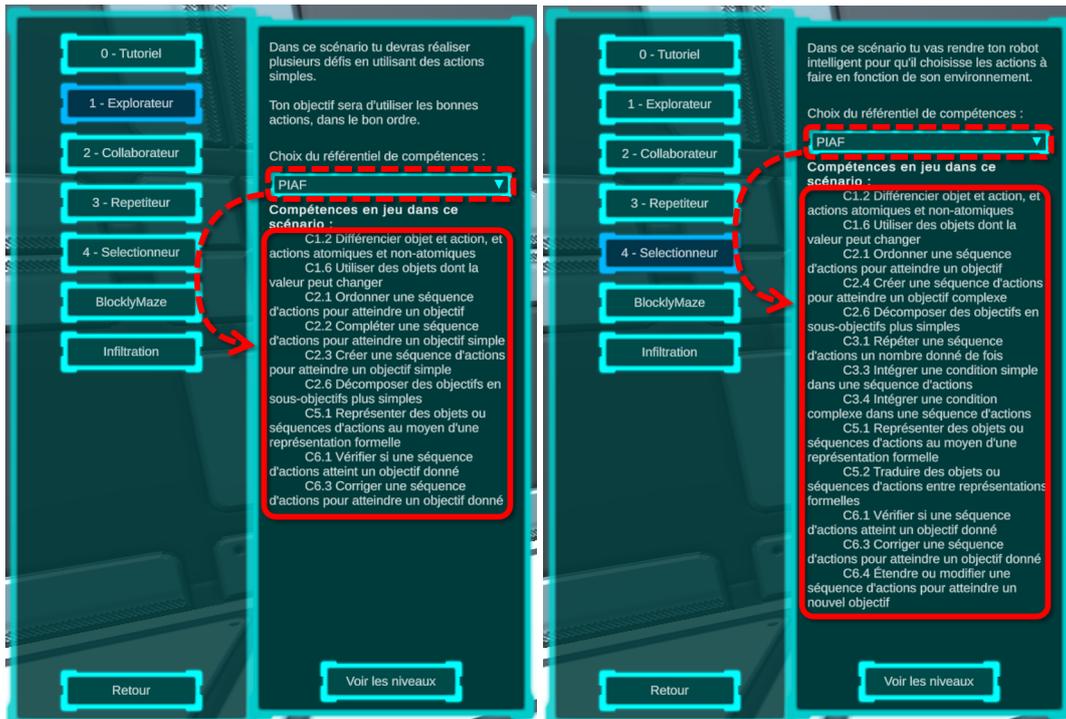
5.1.1. Sélection d'un scénario

Lorsqu'un utilisateur (enseignant ou élève) sélectionne un scénario, la liste des compétences travaillées dans ce scénario lui est proposée (voir Figure 7). Cette liste est construite automatiquement à partir de l'analyse des missions constituant le scénario. L'utilisateur a la possibilité de sélectionner le référentiel de son choix parmi ceux intégrés dans le jeu. Cette analyse a été pensée pour les enseignants afin de leur permettre d'avoir une synthèse rapide des compétences travaillées dans un scénario sans avoir à jouer eux-mêmes à toutes les missions d'un scénario et en déduire les compétences travaillées.

5.1.2. Composition d'un scénario

SPY contient un éditeur de scénario (voir Figure 8). Cet éditeur permet à l'enseignant de consulter, d'éditer et de créer des scénarios. Le panneau de gauche contient l'ensemble des missions de la base de données de SPY, le panneau central présente les informations sur la mission actuellement sélectionnée (dont l'analyse automatique des compétences mobilisées dans cette mission en fonction du référentiel sélectionné), et le panneau de droite affiche l'ensemble des missions intégrées dans le scénario.

L'analyse automatique des compétences est donc une information complémentaire donnée à l'enseignant pour lui permettre de déterminer si la mission sélectionnée est un candidat potentiel pour être intégré dans sa scénarisation pédagogique.



(a) Scénario « Explorateur »

(b) Scénario « Sélectionneur »

Figure 7 : Analyse automatique de deux scénarios à l’aide du référentiel PIAF. Les compétences identifiées pour chaque mission dans l’encadré à trait plein sont calculées dynamiquement par le système en fonction du choix du référentiel sélectionné (encadré en pointillé).

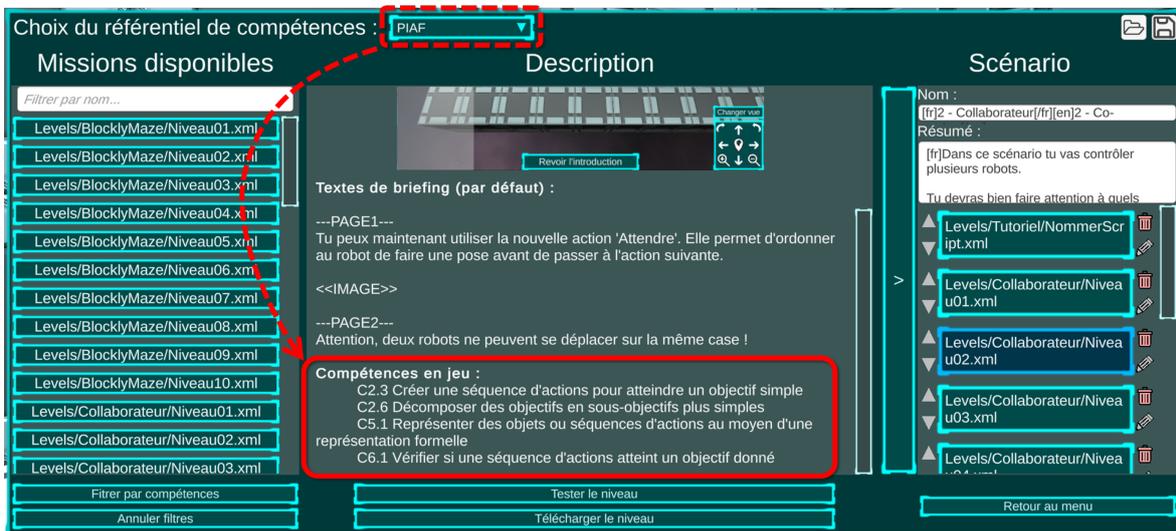


Figure 8 : Vue de l’éditeur de scénario intégré à SPY. La partie encadrée en trait plein est le résultat de l’analyse automatique de la 3ème mission du scénario collaborateur à l’aide du référentiel PIAF (encadré en trait pointillé).

5.1.3. Système de filtre

Comme nous l’avons vu dans la section précédente, l’éditeur de scénario permet aux enseignants d’avoir accès à l’ensemble des missions de la base de données de SPY. Cependant,

identifier les missions permettant de travailler une compétence, ou un ensemble de compétences, n'est pas aisé compte tenu du grand nombre de missions disponibles (près de 60 missions). Nous avons donc détourné l'analyseur automatique de scénario pour créer un filtre de missions par compétences. Cette fonctionnalité permet de sélectionner une ou plusieurs compétences d'un référentiel et de filtrer les missions pour ne proposer à l'enseignant que celles respectant cette contrainte (voir Figure 9).



Figure 9 : Vues de l'éditeur permettant de filtrer les missions par compétences. Dans cet exemple les compétences C3.1 et C3.4 du PIAF ont été sélectionnées. Après application du filtre, seules 5 missions de la base de donnée correspondant à la sélection sont proposées à l'enseignant.

5.2. EXPLOITATION DE L'ANALYSEUR À L'EXTÉRIEUR DE SPY

Nous avons également exploité les résultats de l'analyseur automatique des missions pour étudier l'évolution de la complexité d'un scénario de jeu. Nous avons réalisé ce travail sur deux scénarios de SPY : le scénario « Infiltration » composé de 20 missions et le scénario « BlocklyMaze », une réplique du jeu du même nom⁵, composé de 10 missions. Nous avons utilisé le référentiel du PIAF pour caractériser la dimension didactique des différents scénarios et le référentiel SPY pour caractériser leur dimension ludique. Le cumul des compétences et fonctionnalités ludiques dans chaque mission donne un indicateur de sa complexité (voir Figure 10). 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.* (2017) qui analysent manuellement chaque mission de jeu selon les dimensions ludiques et pédagogiques. Dans notre cas, l'analyse des missions est automatisée. Cette représentation graphique permet ainsi de comparer la complexité de différents scénarios selon les mêmes critères d'analyse. Nous constatons que le scénario Infiltration commence avec un niveau de complexité moindre que BlocklyMaze et se termine avec un niveau de complexité plus élevé. En revanche, la progression de la complexité semble plus régulière dans BlocklyMaze. Cette visualisation est également utile aux concepteurs pour détecter les compétences non traitées dans un scénario ou pour identifier les missions mal équilibrées ou mal positionnées dans le scénario. Par exemple, dans le scénario Infiltration, la compétence C4 n'est traitée que dans la mission 7. Il pourrait donc être intéressant de créer de nouvelles missions mobilisant les compétences

5. BlocklyMaze : <https://blockly.games/maze>, consulté le 08/11/2023

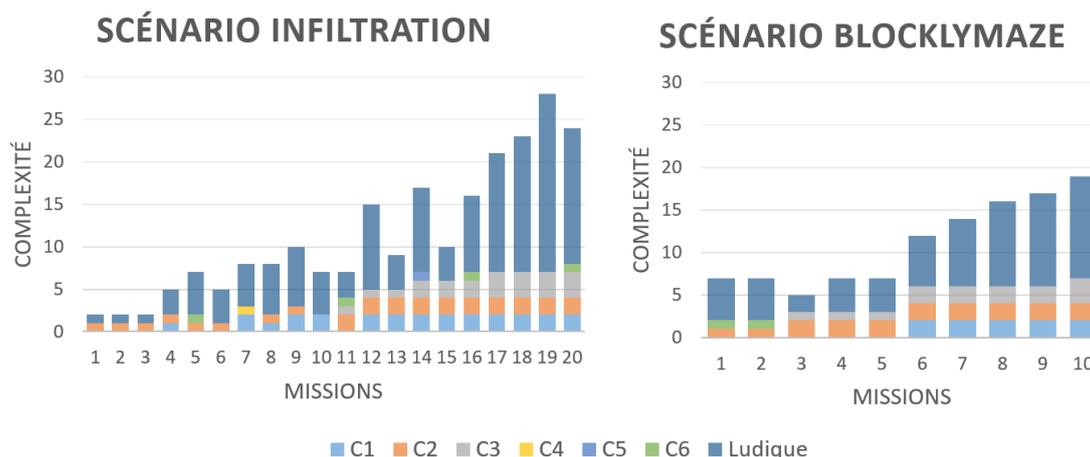


Figure 10 : Analyse des deux scénarios du jeu SPY. Chaque bâton représente un indicateur de la complexité d'une mission obtenue par l'addition du nombre de compétences mobilisées et du nombre de fonctionnalités ludiques impliquées.

C4.2 et C4.3 du PIAF. Cette analyse révèle aussi des pics de complexité (missions 12 et 14 du scénario Infiltration) qui invitent peut-être à étudier ces missions plus en détail et à envisager des ajustements.

6. EXPÉRIMENTATION : CONTEXTE ET PROTOCOLE EXPÉRIMENTAL

Nous avons présenté dans la section 5 comment le module d'analyse automatique a été intégré à SPY pour développer de nouvelles fonctionnalités (analyse d'un scénario, analyse d'une mission, filtrage). Ce module constitue donc une réponse à notre deuxième question de recherche : comment exploiter ce lien [GTC] pour analyser automatiquement les missions du jeu sérieux SPY et en extraire les compétences de la pensée informatique ?

Après un premier temps d'implémentation du module d'analyse automatique sur la base de l'expertise de l'équipe de conception du jeu, il est nécessaire d'interroger la qualité des résultats produits. Afin d'aborder cette question, nous avons conduit une expérimentation en collaboration avec des enseignants en informatique. En effet, si la finalité du module est de contribuer à la problématique générale énoncée dans l'introduction (aider des enseignants peu familiers avec la pensée informatique à repérer, dans des situations de jeu, les compétences mobilisées), l'évaluation de la qualité des résultats produits par l'analyseur automatique de mission nécessite une certaine expertise en informatique et en son enseignement.

6.1. PROFIL DES PARTICIPANTS

Plus précisément, l'expérimentation a été menée avec un groupe de cinq étudiants (une femme et quatre hommes), dans le cadre de la deuxième année d'une formation d'enseignants d'informatique dans le canton de Genève en Suisse. Elle a été organisée dans le cadre d'une séance du séminaire "Recherche de didactique de l'informatique". Cette séance a donc participé à la formation à la recherche des étudiants en leur permettant de vivre une situation de recherche. En raison de leur participation à cette formation, tous les participants partagent des caractéristiques communes :

1. une formation universitaire de niveau Master 2 en informatique ou jugée équivalente ;

2. une année minimum d'enseignement de l'informatique au secondaire 1 (équivalent du collège français);
3. des enseignements d'informatique assurés au secondaire 2 (équivalent du lycée français) depuis la rentrée scolaire 2023.

Pour compléter ce profil, nous avons posé différentes questions aux participants. Tout d'abord, en plus de leur formation initiale en informatique, quatre des cinq participants ont une expérience professionnelle dans le domaine de l'informatique, pour des durées allant de 2 à 25 ans. Concernant leurs expériences d'enseignement, aucun n'a enseigné l'informatique au primaire. Cependant, trois d'entre eux ont entre quelques mois et 5 années d'expérience d'enseignement dans le secondaire (en plus de celle acquise pendant leur formation).

Par rapport à l'artefact SPY sur lequel ils ont eu à travailler durant l'expérimentation :

- tous les participants ont déjà utilisé au moins un environnement de programmation par blocs pour enseigner l'informatique (entre un et trois parmi Scratch, Blockly-games, microbit, code.org, Future Engineer);
- tous les participants estiment que la programmation par bloc est une bonne métaphore pour initier les élèves à la programmation;
- tous les participants signalent utiliser l'approche par compétences pour définir des objectifs d'apprentissage et analyser / concevoir des ressources.

Ces éléments confirment une familiarité des participants avec un environnement du type de SPY et avec l'enseignement de l'informatique. Nous pouvons donc faire l'hypothèse qu'ils présentent un profil pertinent pour évaluer la qualité des résultats produits par l'analyseur de compétences.

6.2. PROTOCOLE EXPÉRIMENTAL

Pour mener cette expérimentation, nous avons conçu un scénario spécifique (*ScenExp*) composé d'une sélection de six missions. Ce scénario joue sur les différentes fonctions et variables présentées dans la section 3.1, afin de permettre l'identification de compétences variées.

Le protocole expérimental a été décomposé en deux phases. La première phase était organisée à distance en modalité asynchrone. Chaque participant avait un travail individuel à réaliser : jouer le scénario *ScenExp* et procéder à une double analyse des missions. La deuxième analyse n'était accessible qu'à l'issue de la première, celle-ci n'étant plus modifiable.

- Première analyse : chaque participant devait décrire les compétences qui lui semblaient mobilisées dans chaque mission avec leurs propres mots. Il leur était demandé de décrire chaque compétence sous la forme « être capable de + verbe action + complément » et de justifier l'identification de cette compétence à partir d'éléments présents dans la mission.
- Seconde analyse : le référentiel PIAF, avec une courte description de chacune des compétences, était fourni aux participants. Ils devaient, pour chaque mission, indiquer la présence ou l'absence de ces compétences et justifier leurs choix. Cette phase avait comme objectif à la fois de faire découvrir le jeu SPY à tous les participants, et également d'introduire le référentiel PIAF afin de créer un espace de réflexion commun pour la phase 2. Les participants ont eu une semaine pour réaliser cette double analyse des missions.

La seconde phase a eu lieu en présentiel, lors d'une session de travail de 3 heures. Les deux auteurs de l'article étaient présents lors de cette session, l'un (*auteur1*) en tant que concepteur de SPY et animateur de la séance, l'autre (*auteur2*) en tant que responsable du

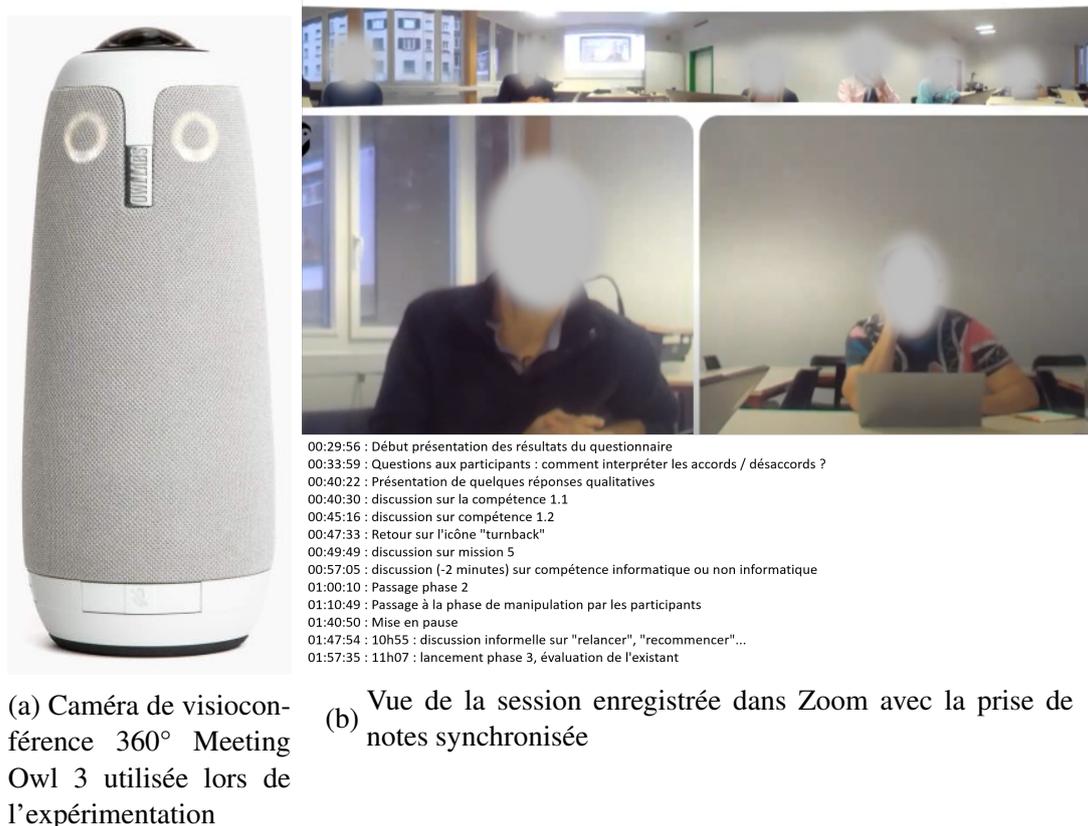


Figure 11 : Dispositif de captation de la session de travail

séminaire recherche et avec une posture plus externe au déroulé de la séance. Des traces audio et vidéo ont été produites à l'aide d'un dispositif de prise de vues 360 (voir Figure 11a) couplé au logiciel de visioconférence Zoom (voir Figure 11b). Le module de discussion de Zoom a été utilisé par *auteur2* comme outil d'annotation. Cette prise de notes en temps réel est ainsi automatiquement synchronisée avec les flux vidéo et audio de l'enregistrement de la séance.

Cette phase a été découpée en plusieurs temps : accueil (5 min), retour collectif sur le travail engagé dans la phase 1 et discussions (30 min), pause (10 min), présentation dans SPY de l'intégration de l'analyseur automatique (10 min), manipulation individuelle de l'éditeur de scénario (45 min), évaluation de l'utilité perçue des fonctionnalités liées à l'analyseur automatique (30 min), discussions informelles et clôture (30 min).

7. RÉSULTATS ET ANALYSES

7.1. PREMIER TEMPS DE LA PHASE 2 : RETOUR COLLECTIF SUR LE TRAVAIL DE LA PHASE 1

Lors du premier temps de discussion avec les participants, nous revenons sur la deuxième partie du questionnaire. Chaque participant devait signaler quelles compétences du PIAF il estimait être mobilisées dans chaque mission. Sur les cinq participants prenant part à l'expérimentation, quatre ont complété le questionnaire proposé pour la phase 1. Le cinquième répondant (R5) a bien joué au scénario proposé et a complété le questionnaire, mais une session expirée a entraîné la perte de ses données. Il disposait tout de même de tous les éléments

lui permettant de participer de manière pertinente à la phase 2.

Pour servir de base à la discussion, nous présentons les résultats de ce questionnaire (les quatre répondants sont notés R1 à R4 et l'analyse produite par le système, nommé Sys) à l'aide du tableau présenté dans la figure 12. Il présente les accords mutuels entre les différents juges sur la présence ou non des différentes compétences du PIAF pour chaque mission. Chaque ligne représente l'accord entre deux juges. Nous considérons que deux juges sont d'accord s'ils ont tous les deux identifié une même compétence sur une même mission ou au contraire s'ils ont jugé une compétence absente sur une même mission. Par exemple, le tableau indique un accord moyen de 0,5 entre R1 et R2 sur la compétence C1.3 car R1 a identifié C1.3 sur les missions 1, 2, 3, 5 et 6, alors que R2 l'a identifiée seulement sur les mission 5 et 6. Ils sont donc en accord sur les missions 4 (absence), 5 (présence) et 6 (présence) et en désaccord sur les missions 1, 2 et 3, soit un accord moyen de 0,5. Les colonnes grisées représentent les compétences reconnues par le système comme étant présentes dans au moins une mission du scénario. La colonne « Moyenne globale » représente l'accord moyen entre deux juges sur l'ensemble des compétences (toutes les colonnes), la colonne « Moyenne + » représente l'accord moyen sur les seules compétences identifiées comme actives par le système (colonnes grisées), et la colonne « Moyenne - » représente l'accord moyen sur les seules compétences identifiées comme absentes par le système (colonnes blanches). Dans ces trois dernières colonnes, les meilleurs accords sont mis en évidence sur fond vert.

		C1.1	C1.2	C1.3	C1.4	C1.5	C1.6	C1.7	C2.1	C2.2	C2.3	C2.4	C2.5	C2.6	C3.1	C3.2	C3.3	C3.4	C4.1	C4.2	C4.3	C5.1	C5.2	C6.1	C6.2	C6.3	C6.4	Moyenne globale	Moyenne +	Moyenne -	
R1	R2	0	0	0,5	1	1	0,5	0,7	0,2	0,8	1	0,8	0,8	0,5	1	1	0,8	1	0,7	0,3	0,3	1	1	1	0	1	1	0,69	0,79	0,56	
R1	R3	1	0	0,8	0	0,5	1	1	0	1	1	0,8	0,7	0,7	1	0,3	0,8	1	0,7	0,8	1	1	0,8	1	0	1	0,3	0,71	0,73	0,67	
R1	R4	1	0,2	0,8	0	0,7	0,5	0,7	0,2	0,5	0,7	0,7	0,7	0,5	1	0,7	1	0,8	0,7	0,3	0,3	0,5	0,7	0,7	0,2	0,2	0,7	0,56	0,59	0,53	
R1	Sys	0	0,5	0,2	1	1	0,3	0,3	0,7	1	1	1	0,5	0,5	1	0,8	0,8	1	1	1	1	1	0,8	0,8	1	1	0,8	0,78	0,82	0,71	
R2	R1	0	0	0,5	1	1	0,5	0,7	0,2	0,8	1	0,8	0,8	0,5	1	1	0,8	1	0,7	0,3	0,3	1	1	1	0	1	1	0,69	0,79	0,56	
R2	R3	0	1	0,3	0	0,5	0,5	0,7	0,8	0,8	1	1	0,8	0,8	1	0,3	1	1	1	0,5	0,3	1	0,8	1	1	1	0,3	0,72	0,86	0,53	
R2	R4	0	0,8	0,3	0	0,7	1	1	1	0,7	0,7	0,8	0,8	0,7	1	0,7	0,8	0,8	1	0,7	0,7	0,5	0,7	0,7	0,8	0,2	0,7	0,68	0,73	0,61	
R2	Sys	1	0,5	0,7	1	1	0,8	0,7	0,2	0,8	1	0,8	0,7	0,7	1	0,8	1	1	0,7	0,3	0,3	0,8	0,8	1	0	1	0,8	0,75	0,83	0,64	
R3	R1	1	0	0,8	0	0,5	1	1	0	1	1	0,8	0,7	0,7	1	0,3	0,8	1	0,7	0,8	1	1	0,8	1	0	1	0,3	0,71	0,73	0,67	
R3	R2	0	1	0,3	0	0,5	0,5	0,7	0,8	0,8	1	1	0,8	0,8	1	0,3	1	1	1	0,5	0,3	1	0,8	1	1	1	0,3	0,72	0,86	0,53	
R3	R4	1	0,8	1	1	0,8	0,5	0,7	0,8	0,5	0,7	0,8	0,7	0,8	1	0,3	0,8	0,8	1	0,5	0,3	0,5	0,8	0,7	0,8	0,2	0,7	0,72	0,72	0,71	
R3	Sys	0	0,5	0	0	0,5	0,3	0,3	0,3	1	1	0,8	0,8	0,8	1	0,2	1	1	0,7	0,8	1	0,8	0,7	1	0	1	0,5	0,62	0,76	0,44	
R4	R1	1	0,2	0,8	0	0,7	0,5	0,7	0,2	0,5	0,7	0,7	0,7	0,5	1	0,7	1	0,8	0,7	0,3	0,3	0,5	0,7	0,7	0,2	0,2	0,7	0,56	0,59	0,53	
R4	R2	0	0,8	0,3	0	0,7	1	1	1	0,7	0,7	0,8	0,8	0,7	1	0,7	0,8	0,8	1	0,7	0,7	0,5	0,7	0,7	0,8	0,2	0,7	0,68	0,73	0,61	
R4	R3	1	0,8	1	1	0,8	0,5	0,7	0,8	0,5	0,7	0,8	0,7	0,8	1	0,3	0,8	0,8	1	0,5	0,3	0,5	0,8	0,7	0,8	0,2	0,7	0,72	0,72	0,71	
R4	Sys	0	0,3	0	0	0,7	0,8	0,7	0,2	0,5	0,7	0,7	0,5	0,7	1	0,8	0,8	0,8	0,7	0,3	0,3	0,7	0,8	0,7	0,2	0,2	0,8	0,53	0,66	0,36	
Sys	R1	0	0,5	0,2	1	1	0,3	0,3	0,7	1	1	1	0,5	0,5	1	0,8	0,8	1	1	1	1	1	0,8	0,8	1	1	1	0,8	0,78	0,82	0,71
Sys	R2	1	0,5	0,7	1	1	0,8	0,7	0,2	0,8	1	0,8	0,7	0,7	1	0,8	1	1	0,7	0,3	0,3	0,8	0,8	1	0	1	0,8	0,75	0,83	0,64	
Sys	R3	0	0,5	0	0	0,5	0,3	0,3	0,3	1	1	0,8	0,8	0,8	1	0,2	1	1	0,7	0,8	1	0,8	0,7	1	0	1	0,5	0,62	0,76	0,44	
Sys	R4	0	0,3	0	0	0,7	0,8	0,7	0,2	0,5	0,7	0,7	0,5	0,7	1	0,8	0,8	0,8	0,7	0,3	0,3	0,7	0,8	0,7	0,2	0,2	0,8	0,53	0,66	0,36	
																										Moyenne	0,68	0,75	0,58		

Figure 12 : Accord mutuel entre les différents juges (système inclus)

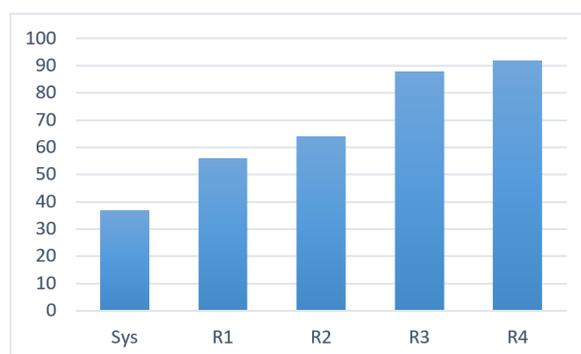


Figure 13 : Nombre de compétences identifiées par chaque juge sur les 6 missions du scénario

Nous remarquons (toujours dans la figure 12) que les répondants 1 et 2 sont plutôt en accord avec le système avec une moyenne globale de 0,78 pour R1 et 0,75 pour R2. Le répondant 3, lui, est plutôt en accord avec les autres juges humains. Le répondant 4 est plutôt en accord avec le répondant 3. Enfin nous remarquons que la moyenne des accords sur la présence des compétences (colonne « Moyenne + ») est plus importante (0,75) que sur l'absence des compétences (colonne « Moyenne - » ; 0,58). Pour comprendre cette tendance nous comparons dans la figure 13 le nombre de compétences identifiées par chaque juge et nous voyons que les juges humains (R1 à R4) ont eu tendance à identifier plus de compétences que le système, ou que le système a eu tendance à éliminer plus de compétences que les juges humains, ce qui explique un désaccord plus fort sur la colonne « Moyenne - ».

L'objectif annoncé aux répondants lors de ce premier échange n'était pas de chercher le consensus à l'issue de la discussion, mais de permettre à chacun d'exprimer son point de vue, de mettre en évidence le côté non homogène des analyses produites et d'identifier des causes de cette non-homogénéité. C'est sur ces éléments qualitatifs que nous revenons maintenant⁶.

Par exemple, la compétence C1.1 (« Être capable de donner des noms à des objets, des actions et des séquences d'actions ») du PIAF n'est pas interprétée de la même manière entre les répondants. Une première interprétation faite par le répondant R1 est de considérer que dans le jeu, tous les blocs sont nommés et que, par conséquent, cette compétence est présente dans toutes les missions puisque les joueurs manipulent des blocs. À l'inverse, R2 indique « *Moi je l'avais compris à l'inverse, pour moi celle-ci j'avais mis qu'elle n'était jamais travaillée dans aucune mission parce que tel que j'ai compris l'intitulé de la compétence, c'est être capable pour l'apprenant de donner des noms à des choses, et ça on ne le fait jamais finalement* ». La règle intégrée dans le système d'analyse est construite selon la même compréhension de la compétence que ce dernier répondant, ce qui explique l'accord de 1 entre R2 et Sys pour la compétence C1.1 en opposition à l'accord de 0 entre les autres répondants et le système sur cette même compétence.

Certaines compétences n'ont tout simplement pas été comprises par certains répondants, comme la compétence C1.2 (« Différencier objet et action, et actions atomiques et non-atomiques », la définition d'action atomique n'étant pas claire pour eux, R3 demande « *ça veut dire quoi action atomique ?* »). Elle est donc absente des compétences identifiées dans les 6 missions, alors même que cette compétence fait partie des compétences identifiées par le système. R2 indique « *Moi je l'ai compris comme ça, de comprendre le fait que chaque bloc finalement est lié à une action, à un truc particulier et de savoir finalement les différencier pour utiliser les bons blocs en fonction de ce qu'on veut faire, donc pour moi c'est vrai que s'en était une qui elle à l'inverse de la une était présente souvent* ». L'interprétation de la compétence pour ce répondant est donc différente de la règle définie pour le système qui cherche à identifier la présence d'actions non-atomiques « Turn back » équivalentes à deux actions atomiques « Turn Left/Right » ou la présence d'actions et de capteurs donnant des informations sur l'état du monde.

Lors des réponses données à la première partie du questionnaire (description des compétences sans s'appuyer sur le PIAF), certaines compétences sont identifiées sans pouvoir être décrites dans le PIAF. Ainsi, le répondant R3 identifie la compétence suivante « *Être capable de se repérer dans l'espace, parce que là faut être capable à l'avance quand même de voir où le bonhomme va se déplacer, savoir s'il doit tourner à gauche pour lui mais du coup ça a rien avoir avec l'informatique* », R2 relance « *Non mais c'est un truc auquel on se heurte, moi je m'y suis heurté en début d'année avec mes élèves où on a fait du turtle [...] la représentation dans l'espace c'était une catastrophe [...] finalement c'est une complexité que j'avais pas anticipée* ». On voit donc ici que même si cette compétence n'est pas jugée

6. Toutes les citations sont issues du verbatim de la séance d'expérimentation.

comme relevant de l'informatique, c'est une compétence qui est travaillée à l'occasion de mises en situation provoquées par l'intention de travailler des compétences en informatique. Sur un tout autre sujet, le répondant R5 identifie une autre compétence : « *Savoir découvrir le fonctionnement de la caméra parce qu'on n'a pas toujours la bonne vue pour savoir ce qu'il faut faire [...] Si on voulait vraiment faire le truc du premier coup juste, il fallait manipuler la caméra pour savoir où il fallait aller* ». Dans ce cas, cette compétence est jugée comme relevant de l'informatique, car elle consiste à prendre des informations sur la situation avant de commencer à programmer. Cette compétence est jugée absente du PIAF et n'est d'ailleurs pas intégrée aux règles présentées dans l'Annexe 8. Pour autant, cette remarque donne une piste pour ajouter une nouvelle règle qui pourrait décrire une facette de la compétence C1.3 (« Identifier les paramètres d'entrée d'une séquence d'actions »). Si une mission donne une vue partielle de la situation, elle invite le joueur à déplacer la caméra pour en comprendre tous les éléments et ainsi identifier tous les paramètres du problème avant de construire sa solution.

Nous voyons donc ici, au travers de ces quelques exemples, que l'interprétation d'un référentiel de compétences pour décrire une tâche donnée est une activité complexe sujette à l'expertise et l'interprétation de chaque enseignant. Pour avoir une compréhension plus précise des éventuels liens entre l'expertise des enseignants, l'analyse qu'ils produisent et l'analyse du système, il conviendrait de poursuivre des expérimentations similaires avec un panel de participants plus important.

7.2. ÉVALUATION DE L'UTILITÉ PERÇUE DES FONCTIONNALITÉS LIÉES À L'ANALYSEUR AUTOMATIQUE

Suite aux échanges sur les résultats du questionnaire, un temps de manipulation libre de l'éditeur de mission a été proposé aux participants. L'objectif de ce temps et des échanges qui ont suivi, était d'évaluer l'utilité perçue des fonctionnalités liées à l'analyseur automatique. Pour aider au recueil des retours des apprenants, nous avons proposé un questionnaire composé de sept questions. Les réponses étaient données à l'aide d'une valeur dans une échelle de Likert à 5 niveaux allant de « 1 - pas du tout » à « 5 - tout à fait ». La progression dans le questionnaire était contrôlée par les animateurs de la séance, les répondants passaient à la question suivante une fois la discussion sur la question courante terminée. Les questions 1, 2 et 3 portaient sur l'évaluation des fonctionnalités déjà intégrées au jeu et les questions 4, 5, 6 et 7 portaient sur des propositions de nouvelles fonctionnalités potentielles. Nous présentons maintenant les réponses à ces différentes questions accompagnées de quelques éléments qualitatifs tirés du verbatim de la séance.

7.2.1. Évaluation de l'utilité perçue des fonctionnalités déjà intégrées

Q1 : Pour vous, la visualisation des compétences au niveau d'un scénario vous semble utile. Tous les répondants (5/5) ont voté 5 (« tout à fait ») :

- R1 : « *Moi je trouve très bien parce que ça permet à l'enseignant de voir quelles compétences il peut directement viser pour l'élève* » ;
- R5 : « *Il peut aussi faire une progression pour vérifier que ce n'est pas toujours les mêmes compétences travaillées* » ;
- R2 : « *La seule chose qui m'a manqué là-dessus, c'est qu'on a l'agrégation complète de tout ce qui a dans le scénario et on n'a pas la séquence, ça aurait été intéressant d'avoir, alors je ne sais pas sous quelle forme le présenter parce que ça pourrait vite être lourd, peut-être sous forme graphique si c'est possible, de voir dans mon scénario la progression des compétences, ça, ça serait un vrai vrai plus* ».

La remarque du répondant R2 est particulièrement intéressante car elle anticipe directement la question 7 du questionnaire.

Q2 : Pour vous, pouvoir filtrer les missions par compétence vous semble utile. Tous les répondants (5/5) ont voté 5 (« tout à fait ») :

- R2 : « *Quand on veut utiliser quelque chose ou [...] si on veut s'inspirer de quelque chose, pouvoir filtrer c'est hyper intéressant* » ;
- R5 : « *Si on repère que quelque chose n'a pas été acquis par la plupart des élèves, on peut proposer des exercices complémentaires liés à ça* », R1 surenchérit « *exactement* ».

On voit donc ici que la fonction de filtre semble utile pour : (1) rechercher des missions existantes qui pourront servir d'inspiration à la création de nouvelles missions ; et (2) sélectionner un ensemble de missions complémentaires à proposer aux élèves pour les faire travailler une compétence précise jugée non encore acquise. On notera dans la remarque de R5 que l'adaptation envisagée est pensée au niveau du groupe classe (« [...] la plupart des élèves [...] ») et non pas à un niveau individuel.

Q3 : Pour vous, l'analyse automatique des compétences d'une mission vous semble utile. Trois répondants ont voté 4 et deux ont voté 5 (« tout à fait ») :

- R2 : « *C'est cool* », R1 surenchérit « *Je pense que c'est moins pertinent, mais c'est cool [rires]* » ;
- R5 : « *Moi je mets 4 parce que ça dépend vraiment de comment est déterminé tout ça, par exemple certaines choses je les trouve détectées de manière assez généreuse* », R2 complète « *Oui, c'est super utile, mais il faut probablement affiner la manière de les détecter* ».

La remarque de R5 est intéressante car il juge le système « généreux ». En d'autres termes, il considère que certaines compétences du PIAF ont été identifiées par le système et ne le devraient peut-être pas. Dans le même sens, R2 suggère de réviser les règles afin de les rendre plus précises. Pour autant, nous avons vu dans la figure 13 que le système est le plus strict des juges avec 37 occurrences de compétences identifiées contre 64 pour R2 et plus de 90 pour R4. Les règles intégrées dans le système d'analyse automatique sont donc plus restrictives que les règles implicites de chaque participant. Indépendamment donc de la réelle finesse d'une analyse, deux questions sous-jacentes sont mises en évidence par les commentaires des répondants.

La première porte sur la confiance portée sur l'analyse et la capacité à expliquer les raisons qui mènent le système à produire un résultat. R2 indique à ce sujet « *Si l'analyse on peut lui faire confiance, parce que les règles sont bien posées, là ça a une vraie valeur, parce que c'est vrai qu'on n'a pas tous les critères en tête quand on crée une mission [...] je pense que c'est une vraie valeur ajoutée pour autant qu'on puisse lui accorder une confiance raisonnable* ».

La seconde porte sur le positionnement des compétences entre elles au sein d'une même mission. R5 propose « *On pourrait avoir des référentiels qui pourraient être structurés autrement et pour lesquels ce n'est pas juste vrai ou faux [présence/absence de la compétence] mais où il y aurait un niveau pour indiquer si elle est plus ou moins présente [...] Indiquer la compétence qui est vraiment centrale à la résolution du problème et une qui est vraiment annexe, et ça on ne le voit pas [...] elles sont toutes à plat* ».

Ces trois premières questions portaient donc sur les fonctionnalités actuellement intégrées dans le jeu SPY. Les évaluations données, ainsi que les remarques formulées, montrent que ces fonctionnalités sont unanimement perçues comme utiles par les cinq répondants. La dernière proposition est légèrement moins plébiscitée que les deux premières mais des pistes d'amélioration intéressantes sont dégagées (explication des règles et positionnement

des compétences entre elles).

7.2.2. Évaluation de l'utilité *a priori* de nouvelles fonctionnalités

Les quatre questions suivantes portent maintenant sur des pistes de développements à venir dans SPY.

Q4 : Actuellement, l'analyseur de compétences par mission est disponible dans l'éditeur de scénario. Disposer également de l'analyseur de compétences dans l'éditeur de mission vous semblerait utile. Quatre répondants ont voté 4 et un a voté 5 (« tout à fait »).

R2 indique à ce sujet « *À chaque fois qu'on ajoute ou qu'on enlève un truc qu'il actualise [les compétences], ouais ça, ça serait un vrai vrai plus* ». Une discussion s'engage sur l'usage *a priori* d'un tel dispositif et la nécessité pour l'enseignant de garder le contrôle. L'analyseur devrait rester un outil d'aide à la validation de ce qui est pensé par l'enseignant pour confirmer l'intention initiale ou faire apparaître d'éventuels écarts et inviter l'enseignant à retravailler sa mission. R2 synthétise ainsi « *Pour ne pas tomber dans un biais où on va aller un peu au hasard, si on sait qu'on veut faire un scénario qui développe telle compétence et qu'on a moyen de se rafraîchir les idées en affichant les règles, pour cette compétence [...] je vais pouvoir imaginer mon scénario en sachant finalement ce que j'inclus dedans* ».

Q5 : Actuellement, le résultat de l'analyse des missions est présenté comme une liste de compétences. Vous semblerait-il utile de pouvoir disposer d'informations expliquant pourquoi telles ou telles compétences ont été identifiées. Deux répondants ont voté 4 et trois ont voté 5 (« tout à fait »).

Cette question fait directement écho aux réflexions qui ont émergé pour la question 3 sur la confiance portée à l'analyseur automatique et aux règles sous-jacentes. Ces résultats confirment la pertinence d'engager des travaux sur l'explicabilité à l'enseignant des décisions du système.

Q6 : Vous semblerait-il utile dans l'éditeur de mission de disposer d'une fonctionnalité capable de suggérer des modifications à apporter à la mission pour pouvoir travailler une compétence visée. Deux répondants ont voté 2, un répondant a voté 3 et les deux derniers ont voté 5.

Cette proposition est la moins bien notée par les répondants. Elle est jugée moins prioritaire que les propositions précédentes.

Q7 : Actuellement, l'analyse des compétences d'un scénario est une simple agrégation des compétences de chaque mission. Disposer d'une analyse plus fine montrant les compétences utilisées dans chaque mission vous semblerait-elle utile. Associée à cette question, l'illustration de la figure 10 était présentée aux répondants. Tous les répondants (5/5) ont voté 5 (« tout à fait »).

L'évaluation de cette proposition est cohérente avec les retours formulés par le répondant R2 dans la question Q1.

7.2.3. Retour sur les discussions informelles

La dernière étape de notre protocole laissait la place à un temps de discussion informel autour du jeu. Ce temps-là a permis de mettre en évidence des limites et des potentialités d'usages de la ressource. Nous en reprenons quelques éléments marquants dans les lignes suivantes.

Concernant les limites, la critique principale a porté sur des choix de développement des interfaces du jeu, en relevant des améliorations possibles d'interactions. R5 note « *Dans les*

moins c'est l'UI, il y a plein d'éléments d'UI qui posent problèmes, les choses liées fonctionnellement ne sont pas toujours liées visuellement ». Ici R5 fait référence à la zone d'édition apparaissant à la gauche de l'écran de jeu (voir Figure 1) et au panneau d'exécution apparaissant sur la droite de l'écran de jeu (voir Figure 3). C'est deux interfaces sont, pour lui, liées fonctionnellement mais ne sont pas visuellement car disposées à l'opposée l'une de l'autre. Pour autant ce choix n'était pas innocent. Il permet d'une part de différencier le programme édité, du programme chargé en mémoire par le robot (le programme exécuté) ; ces deux zones d'affichage distinctes ont été pensées dans ce sens. D'autre part, différencier la zone d'édition du contexte d'exécution permet d'allouer autant de contextes d'exécution que de robots dans la mission. Ainsi, comme nous le montrons dans la figure 3, un même programme (défini dans une et une seule zone d'édition) peut être exécuté par deux robots différents où chaque robot a son propre programme chargé en mémoire et l'exécute indépendamment. Ces choix sont, encore une fois, liés aux savoirs à enseigner. Pour autant, dans cet exemple relevé par R5, ces choix ne sont pas lisibles par l'enseignant et méritent donc d'être retravaillés.

Concernant les potentialités, nous en avons identifié de deux ordres : l'exploitation des fonctionnalités en place et le détournement de fonctionnalités à des fins pédagogiques.

Concernant l'exploitation des fonctionnalités en place, la diversité des objets de jeu interactifs est notée. R5 indique « *Clairement par rapport à un truc comme Blockly, il y a déjà des plus comme le fait qu'on peut programmer différents objets qui interagissent voire de manipuler des terminaux [pour ouvrir des portes] ou des trucs comme ça, là il y a de quoi faire plus donc je trouve ça vraiment intéressant* ». R2 complète en insistant sur la dimension ludique de certains scénarios : « *Moi, j'avais regardé le scénario Infiltration qui a une vingtaine de missions qui sont liées par une histoire finalement avec un vrai récit et je pense que, en créant des scénarios comme ça, il y a une vraie possibilité [...] comme tu dis, il y a beaucoup d'interactions possibles dès qu'on met deux robots, dès qu'on rajoute des drones, des objets manipulables, des pièces à ramasser, des terminaux, etc. Il y a beaucoup de choses finalement faisables [...] et le gros plus de cette plateforme c'est de jouer là-dessus, la différenciation* ».

Concernant les détournements à des fins pédagogiques, une première idée est formulée par R2 : « *Moi, il y a un truc que je me suis tout de suite dit quand j'ai vu des scénarios avec deux robots et deux scripts différents, je me suis dit là ce qui serait génial c'est la coopération, deux élèves se connectent sur la même session et chacun programme son robot et ils doivent discuter ensemble [...] pour se dire, ok, attend, moi mon robot il doit faire trois étapes pour ouvrir la porte donc toi il faut que tu attendes trois fois. Dès que j'ai vu deux robots avec deux scripts distincts je me suis dit mais là il y a un potentiel avec les élèves qui est juste incroyable* ». Une discussion s'en est suivie sur la complexité de développer une telle fonctionnalité. Face à cette difficulté technique, R2 propose alors une solution originale « *après je pense qu'on peut y arriver de manière un peu détournée en mettant deux élèves sur le même ordinateur et que chacun programme son robot* ».

Un second détournement concerne l'éditeur de mission par les élèves. Cet éditeur, initialement conçu pour les enseignants afin de leur permettre d'ajouter des missions originales à leur scénario, pourrait être manipulé par les élèves afin qu'ils créent leurs propres missions, tentent de les résoudre et les proposent à leurs camarades de classe. Ce type d'activité pourrait être dirigé en demandant aux élèves de créer une mission qui nécessite de mobiliser une compétence précise. Dans ce cas, la tâche demandée à l'élève serait au-delà de la simple maîtrise de la compétence mais viserait à comprendre ce qu'est une situation qui nécessite la manipulation de la compétence. Le répondant R2 synthétise ces échanges en scénarisant ce détournement de l'usage de l'éditeur de mission sur les différents niveaux scolaires :

« L'usage classique de jouer des missions au niveau du cycle [collège français], et au niveau du lycée aller sur de la conception de missions là ça pourrait être plus intéressant effectivement ».

Enfin, dans la continuité des discussions portant sur le détournement de l'éditeur de mission, un dernier détournement est évoqué sur l'exploitation des fichiers XML décrivant les missions de jeu. Présenter ces fichiers aux élèves pourrait permettre d'aborder le thème des langages de balisage pour décrire une ressource (dans notre cas, une mission de jeu) et comprendre que l'éditeur de mission n'est qu'une interface permettant de créer ces fichiers qui peuvent être ouverts et modifiés avec un simple éditeur de texte. Là encore, le répondant R2 synthétise les échanges : *« Ça, les élèves adorent, d'aller taper du code finalement et de voir le résultat dans un produit fini, qui est là qui est bien léché je veux dire et de voir que simplement en allant modifier quelques trucs dans un fichier texte comme ça et bien on a un impact direct sur un jeu, ça je pense que ça devrait plaire ».*

8. 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 le formalisme générique GTC basé sur la structuration du langage de balisage XML pour décrire des compétences par combinaison de fonctionnalités de jeu (question de recherche 1). Nous avons ensuite exploité GTC pour proposer une caractérisation des 21 compétences PIAF présentes dans SPY, des 5 niveaux de la compétence 3.4 du CRCN et des 35 fonctionnalités ludiques de SPY. L'ensemble des résultats est présenté dans Muratet (2023b). Ainsi, nous avons montré que GTC 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. Dans ce travail de recherche, les règles spécifiques à SPY ont été conçues par les chercheurs impliqués sur le projet. Le résultat n'est donc pas l'ensemble des règles proprement dit mais le formalisme générique permettant de décrire ces règles : GTC. Un travail reste à engager pour réviser les règles proposées pour SPY et les soumettre à un processus collectif de validation impliquant des enseignants. Ceci devrait également contribuer à mieux expliquer les décisions du système et donc à augmenter la confiance accordée par les enseignants aux informations fournies par celui-ci.

Aussi, nous avons montré comment ces caractérisations peuvent être exploitées pour construire des indicateurs sur des missions existantes et fournir aux enseignants et concepteurs des informations micros sur les compétences en jeu dans une mission, ou des informations macros sur l'évolution de la complexité d'un scénario de jeu (question de recherche 2). Ces résultats ont été intégrés au jeu SPY à travers trois fonctionnalités : pour l'analyse d'une mission dans l'éditeur de scénario ; pour l'analyse d'un scénario par l'agrégation des compétences de chaque mission composant le scénario ; et par le développement d'une fonction de filtrage permettant de filtrer les missions par compétences dans l'éditeur de scénario.

Ces trois fonctionnalités ont été étudiées lors d'une expérimentation menée avec un groupe d'enseignants en formation à l'Université de Genève. Nous avons pu montrer que l'appropriation d'un référentiel de compétences tel que le PIAF reste complexe, même pour des enseignants formés à la discipline informatique. L'accord entre les différents juges pour identifier les compétences mobilisées dans une sélection de missions du jeu SPY reste hétérogène. Les trois fonctionnalités intégrées au jeu ont été jugées très utiles par les participants à l'expérimentation. L'expérimentation a permis également d'identifier des détournements possibles de certaines fonctionnalités du jeu à des fins pédagogiques, comme la manipulation de l'éditeur de mission par les élèves ou la présentation aux élèves des fichiers XML

décrivant les missions.

Ces premiers retours sont encourageants. Ils permettent d’identifier plusieurs pistes de recherche en rapport avec la contribution de cet article :

- quel passage de l’utilité perçue *a priori* à une utilité vérifiée en contexte réel d’utilisation ?
- quels effets du profil des utilisateurs enquêtés, et en particulier de leurs compétences en informatique ?
- comment favoriser la confiance des utilisateurs dans les informations données par le système notamment en travaillant l’explicabilité des décisions prises et des règles sous-jacentes ?

Plus largement se pose la question de la transférabilité de la démarche d’analyse des compétences à partir des fonctionnalités de jeu et notamment des règles proposées dans d’autres environnements dédiés à l’apprentissage de la pensée informatique.

Enfin, SPY s’inscrit dans une démarche de partage à la fois pour son usage (*free to play*), son code source (*open source*) et ses traces d’interaction (*open data*) (Muratet, 2023c). Les scénarios pré-construits sont utilisables en l’état par des enseignants qui souhaiteraient faire travailler leurs élèves sur les compétences de la pensée informatique. Les deux éditeurs inclus (éditeur de scénario et éditeur de mission) permettent aux enseignants de personnaliser le jeu, voire de créer leurs propres missions et scénarios. Nous avons vu également que l’éditeur de mission pourrait être avantageusement détourné pour le transformer en une ressource pédagogique à destination des élèves en les invitant à créer leurs propres missions. Concernant les chercheurs, l’ouverture de son code source et de ses données ouvre la voie au développement de nouvelles fonctionnalités afin de tester leurs hypothèses en s’appuyant sur des observations en contexte de laboratoire ou écologique et en exploitant les données ouvertes du jeu.

REMERCIEMENTS

Cette recherche a été menée dans le cadre du projet ANR-18-CE38-0008 (IECARE) financé par l’Agence Nationale de la Recherche (ANR).

RÉFÉRENCES

- Alayrangues, S., Peltier, S., et Signac, L. (2017). Informatique débranchée : construire sa pensée informatique sans ordinateur. *Actes du colloque Mathématiques en Cycle 3 IREM de Poitiers*, (p. 216-226). <https://hal.science/hal-01868132>
- Balacheff, N. (1994). La transposition informatique, un nouveau problème pour la didactique. Dans M. Artigue, R. Gras, C. Laborde et P. Tavnnot (dir.), *Actes du colloque “Vingt ans de didactique des mathématiques en France”* (p. 364-370). La Pensée Sauvage. <https://telearn.archives-ouvertes.fr/hal-00190646>
- Baron, G.-L., Drot-Delange, B., Grandbastien, M., et Tort, F. (2014). Computer science education in french secondary schools : historical and didactical perspectives. *ACM Trans. Comput. Educ.*, 14(2). <https://doi.org/10.1145/2602486>
- Bonnat, C., Sanchez, E., Paukovics, E., et Kramar, N. (2023). Didactic transposition and learning game design. Towards a ludicization model for school visits in museums [Series Title : Transdisciplinary Perspectives in Educational Research]. Dans F. Ligozat, K. Klette et J. Almqvist (dir.), *Didactics in a Changing World* (p. 199-215, T. 6). Springer International Publishing. https://doi.org/10.1007/978-3-031-20810-2_12

- Brousseau, G. (1986). Fondements et méthodes de la didactique des mathématiques. *Recherches en didactique des mathématiques*, 7(2), 33-115.
- Brousseau, G. (1990). Le contrat didactique : le milieu. *Recherches en didactique des mathématiques*, 9(3), 309-336. <https://hal.science/hal-00686012/>
- Bruillard, É. (1997). *Les machines à enseigner*. Editions Hermès.
- Carron, T., Muratet, M., Marne, B., et Yessad, A. (2017). Analyser et représenter la progression de la difficulté d'un jeu sérieux du point de vue ludique et pédagogique. *EIAH 2017*. <https://hal.science/hal-01515753>
- Komis, V., et Misirli, A. (2011). Robotique pédagogique et concepts préliminaires de la programmation à l'école maternelle : une étude de cas basée sur le jouet programmable Bee-Bot. Dans G.-L. Baron, É. Bruillard et V. Komis (dir.), *Actes du colloque international DIDAPRO 4 - Dida&Stic* (p. 271-281). New Technologies Editions. <https://edutice.hal.science/edutice-00676143>
- Kradolfer, S., Dubois, S., Riedo, F., Mondada, F., et Fassa, F. (2014). A sociological contribution to understanding the use of robots in schools : the Thymio robot. Dans M. Beetz, B. Johnston et M.-A. Williams (dir.), *Actes de la Conférence Internationale Social Robotics (ICSR)* (p. 217-228). Springer International Publishing.
- Lindberg, R. S. N., Laine, T. H., et Haaranen, L. (2019). 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. <https://doi.org/https://doi.org/10.1111/bjet.12685>
- Margolinas, C. (1992). Eléments pour l'analyse du rôle du maître : les phases de conclusion. *Recherches en didactique des mathématiques*, 12(1), 113-158.
- Mariotti, M. A., et Maracci, M. (2012). Resources for the teacher from a semiotic mediation perspective. Dans G. Gueudet, B. Pepin et L. Trouche (dir.), *From text to 'lived' resources : mathematics curriculum materials and teacher development* (p. 59-75). Springer Netherlands.
- Marne, B., Wisdom, J., Huynh-Kim-Bang, B., et Labat, J.-M. (2012). The six facets of serious game design : a methodology enhanced by our design pattern library. *Actes de la 7ème European Conference on Technology Enhanced Learning (EC-TEL 2012)*, (p. 208-221). https://doi.org/10.1007/978-3-642-33263-0_17
- Miljanovic, M. A., et Bradbury, J. S. (2018). A review of serious games for programming. Dans S. Göbel, A. Garcia-Agundez, T. Tregel, M. Ma, J. Baalsrud Hauge, M. Oliveira, T. Marsh et P. Caserman (dir.), *Actes de la conférence Serious Games* (p. 204-216). Springer International Publishing.
- Muratet, M. (2023a). Complete and commented level model in XML format [[Online; accessed June 19, 2023]]. <https://github.com/Mocahteam/SPY/blob/master/Doc/LevelModel.xml>
- Muratet, M. (2023b). Description of PIAF skills using the SPY game features [[Online; accessed June 19, 2023]]. <https://github.com/Mocahteam/SPY/blob/master/Assets/StreamingAssets/Competencies/competenciesReferential.json>
- Muratet, M. (2023c). SPY : un jeu sérieux partagé pour étudier l'apprentissage de la pensée informatique. Dans J. Broisin, C. Declercq, C. Fluckiger, Y. Parmentier, Y. Peter et Y. Secq (dir.), *Atelier "Apprendre la Pensée Informatique de la Maternelle à l'Université", dans le cadre de la conférence Environnements Informatiques pour l'Apprentissage Humain (EIAH)* (p. 33-40). <https://hal.science/hal-04144205>
- Parmentier, Y., Reuter, R., Higuette, S., Kataja, L., Kreis, Y., Duflot-Kremer, M., Laduron, C., Meyers, C., Busana, G., Weinberger, A., et Denis, B. (2020). PIAF : developing computational and algorithmic thinking in fundamental education. *EdMedia+ Inno-*

- vate Learning, Association for the Advancement of Computing in Education (AACE), 315-322.
- PIAF Project. (2021). The complete list of the 26 PIAF skills and their description [[Online; accessed June 19, 2023]]. <https://piaf.loria.fr/wp-content/uploads/2021/09/PIAF-Referential-of-Competencies-Description-and-Examples.pdf>
- Poirson, B. (2020). *Appuis et obstacles à l'apprentissage de la pensée informatique à l'école primaire* [mém. de mast., INSPÉ de Franche-Comté (Vesoul); Université de Franche-Comté (UFC)]. <https://univ-fcomte.hal.science/hal-02958874>
- Rabardel, P. (1995). *Les hommes et les technologies. Approche cognitive des instruments contemporains*. Armand Colin.
- Rabardel, P. (1999). Eléments pour une approche instrumentale en didactique des mathématiques. *Actes de la 10e école d'été de didactique des mathématiques, 18 (21)*, 203-213.
- Saddoug, H., Rahimian, A., Marne, B., Muratet, M., Sehaba, K., et Jolivet, S. (2022). Review of the adaptability of a set of learning games meant for teaching computational thinking or programming in France. *Special Session on Gamification on Computer Programming Learning, 1*, 562-569. <https://doi.org/10.5220/0011126400003182>
- Sigayret, K., Tricot, A., et Blanc, N. (2021). Pensée informatique et activités de programmation : quels outils pour enseigner et évaluer ? Dans J. Broisin, C. Declercq, C. Fluckiger, Y. Parmentier, Y. Peter et Y. Secq (dir.), *Actes de l'atelier "Apprendre la Pensée Informatique de la Maternelle à l'Université", dans le cadre de la conférence Environnements Informatiques pour l'Apprentissage Humain (EIAH)* (p. 68-75). <https://hal.science/hal-03241689>
- Spach, M. (2019). Activités robotiques à l'école : approches de pratiques d'enseignement et effets sur les apprentissages : *Recherches en didactiques, N° 28(2)*, 68-87. <https://doi.org/10.3917/rdid.028.0068>
- Wing, J. M. (2006). Computational thinking. *Communications of the ACM, 49(3)*, 33-35.

ANNEXES

RÉSUMÉ DES COMPÉTENCES ENCODÉES

Nous présentons dans cette annexe un résumé des compétences du PIAF encodées. Nous rappelons pour chaque compétence, son intitulé, une mise en contexte de cette compétence dans SPY, ainsi que la description de la compétence à l'aide de GTC sous une forme encadrée. Pour simplifier la description des compétences dans cette annexe, le tableau 5 répertorie les fonctions récurrentes utilisées.

C1.1 Nommer des objets et séquence d'actions : si une mission contient au moins une zone d'édition nommable.

$$C1.1 := ZEN() \geq 1$$

C1.2 Différencier objet et action, et actions atomiques et non-atomiques : si une mission contient une action non-atomique (l'action « Turn back » (IBQ_1)) qui peut être décomposée en deux actions atomiques (les actions « Turn left » (IBQ_2) ou « Turn right » (IBQ_3)) ou si une mission contient des actions (IEQ_1) et des capteurs (IEQ_2) qui illustrent les concepts d'expressions et d'instructions.

$$C1.2 := (IBQ_1(\text{TurnBack}, 1) \&\& (IBQ_2(\text{TurnLeft}, 2) \parallel IBQ_3(\text{TurnRight}, 2))) \parallel (IEQ_1(\{\text{AllActions}\}, 1) \&\& IEQ_2(\{\text{AllCaptors}\}, 1))$$

C1.3 Identifier les paramètres d'entrée d'une séquence d'actions :

Tableau 5 : fonctions récurrentes

Fonction	Description
$ZEN()$	nombre de Z one d' E dition N ommable $\rightarrow Card(script\ editMode = 2)$
$ZEC()$	nombre de Z one d' E dition C orrecte $\rightarrow Card(script\ type = 0)$
$ZENO()$	nombre de Z one d' E dition N on O ptimale $\rightarrow Card(script\ type = 1)$
$ZEB()$	nombre de Z one d' E dition B oguée $\rightarrow Card(script\ type = 2)$.
$IBQ(B, Q)$	vrai si la mission In clut au moins un B loc B en Q uantité Q , faux sinon $\rightarrow Card((blockLimit\ blockType = B) \cap (blockLimit\ limit \geq Q)) \geq 1$
$IEQ(S, Q)$	vrai si la mission In clut au moins un item de l' E nsemble S en Q uantité Q , faux sinon $\rightarrow Card((blockLimit\ blockType\ isIncludedIn\ S) \cap (blockLimit\ limit \geq Q)) \geq 1$
$DD()$	vrai si le glisser/déposer (D rag& D rop) est activé, faux sinon $\rightarrow Card(dragdropDisabled) = 0$

Aucune mécanique de jeu pour cette compétence.

C1.4 Décrire le résultat d'une séquence d'actions :

Aucune mécanique de jeu pour cette compétence.

C1.5 Prédire le résultat d'une séquence d'actions : si une mission contient une sentinelle avec un programme défini.

$C1.5 := Card(script\ outputLine\ sameValue\ guard\ inputLine \cap script\ hasChild) \geq 1$

C1.6 Utiliser des objets dont la valeur peut changer : si une mission contient une structure de contrôle conditionnelle (IEQ_1) et des capteurs (IEQ_2) dont la valeur varie en fonction du contexte ou si une mission permet de contrôler (IBQ) l'état d'une porte ($R1$).

$R1 \rightarrow Card(door\ slotId\ sameValue\ slot\ slotId) \geq 1$
 $C1.6 := (IEQ_1(\{While, IfThen, IfElse\}, 1) \&\& IEQ_2(\{AllCaptors\}, 1)) \parallel (IBQ(Activate, 1) \&\& R1)$

C1.7 Reconnaître, parmi des objets et séquences d'actions connus, lesquels peuvent être utilisés pour atteindre un nouvel objectif : si une mission propose seulement des zones d'édition nommables ($R1$) dont au moins une contient une solution pré-construite correcte ($R2$) que le joueur doit reconnaître et ne peut modifier (DD).

$R1 \rightarrow ZEN() = Card(script)$
 $R2 \rightarrow Card((script\ type = 0) \cap (script\ hasChild)) \geq 1$
 $C1.7 := R1 \&\& R2 \&\& !DD()$

C2.1 Ordonner une séquence d'actions pour atteindre un objectif : si une mission permet au joueur de combiner des actions (DD) en quantité limitée ($R1$) et avec au moins une action disponible dans l'inventaire (IEQ).

$R1 \rightarrow Card(blockLimit\ limit = -1) = 0$
 $C2.1 := DD() \&\& R1 \&\& IEQ(\{AllActions\}, 1)$

C2.2 Compléter une séquence d'actions pour atteindre un objectif simple : si une mission permet au joueur de combiner (DD) seulement des actions (IEQ_1) sans structures de contrôle (IEQ_2) et en fournissant seulement des zones d'édition pré-construites mais boguées ($R1$) que le joueur doit compléter.

$$R1 \rightarrow Card((script\ type = 2) \cap (script\ hasChild)) = Card(script)$$

$$C2.2 := DD() \ \&\& \ IEQ_1(\{AllActions\}, 1) \ \&\& \ !IEQ_2(\{ControlList\}, 1) \ \&\& \ R1$$

C2.3 Créer une séquence d’actions pour atteindre un objectif simple : si une mission permet au joueur de combiner (*DD*) seulement des actions (*IEQ₁*) sans structures de contrôle (*IEQ₂*).

$$C2.3 := DD() \ \&\& \ IEQ_1(\{AllActions\}, 1) \ \&\& \ !IEQ_2(\{ControlList\}, 1)$$

C2.4 Créer une séquence d’actions pour atteindre un objectif complexe : si une mission permet au joueur de combiner (*DD*) des actions (*IEQ₁*) et des structures de contrôle (*IEQ₂*).

$$C2.4 := DD() \ \&\& \ IEQ_1(\{AllActions\}, 1) \ \&\& \ IEQ_2(\{ControlList\}, 1)$$

C2.5 Combiner des séquences d’actions pour atteindre un objectif : Aucune mécanique de jeu pour cette compétence.

Aucune mécanique de jeu pour cette compétence.

C2.6 Décomposer des objectifs en sous-objectifs plus simples : si une mission ne limite pas le nombre d’exécution à 1.

$$C2.6 := Card(executionLimit\ amount = 1) = 0$$

C3.1 Répéter une séquence d’actions un nombre donné de fois : si une mission permet au joueur de combiner (*DD*) des boucles « Répéter n fois » (*IBQ*).

$$C3.1 := DD() \ \&\& \ IBQ(ForLoop, 1)$$

C3.2 Répéter une séquence d’actions jusqu’à ce qu’un objectif soit atteint : si une mission permet au joueur de combiner (*DD*) des boucles « Tant que » (*IBQ*) avec des capteurs (*IEQ₁*) et sans opérateurs (*IEQ₂*).

$$C3.2 := DD() \ \&\& \ IBQ(WhileLoop, 1) \ \&\& \ IEQ_1(\{AllCaptors\}, 1) \ \&\& \ !IEQ_2(\{OperatorList\}, 1)$$

C3.3 Intégrer une condition simple dans une séquence d’actions : si une mission permet au joueur de combiner (*DD*) la structure de contrôle « Si Alors » (*IEQ₁*) avec des capteurs (*IEQ₂*) et sans opérateurs (*IEQ₃*).

$$C3.3 := DD() \ \&\& \ IEQ_1(\{IfThen, IfElse\}, 1) \ \&\& \ IEQ_2(\{AllCaptors\}, 1) \ \&\& \ !IEQ_3(\{OperatorList\}, 1)$$

C3.4 Intégrer une condition complexe dans une séquence d’actions : si une mission permet au joueur de combiner (*DD*) des structures de contrôle conditionnelles (*IEQ₁*) avec des capteurs (*IEQ₂*) et des opérateurs (*IEQ₃*).

$$C3.3 := DD() \ \&\& \ IEQ_1(\{While, IfThen, IfElse\}, 1) \ \&\& \ IEQ_2(\{AllCaptors\}, 1) \ \&\& \ IEQ_3(\{OperatorList\}, 1)$$

C4.1 Comparer deux objets selon un critère donné : Aucune mécanique de jeu pour cette compétence.

Aucune mécanique de jeu pour cette compétence.

C4.2 Comparer deux séquences d’actions selon un critère donné : si une mission ne contient que des zones d’édition pré-construites nommables (*R1*) avec au moins une solution correcte et d’autres solutions non-optimales (*R2*) que le joueur ne peut modifier (*DD*).

$$R1 \rightarrow ZEN() = Card(script)$$

$$R2 \rightarrow ZEC() \geq 1 \ \&\& \ (ZEC() + ZENO()) = Card(script)$$

$$C4.2 := R1 \ \&\& \ R2 \ \&\& \ !DD()$$

C4.3 Améliorer une séquence d’actions par rapport à un critère donné : si une mission ne contient que des solutions pré-construites non-optimales ($R1$) que le joueur doit améliorer (DD).

$$\begin{aligned} R1 &\rightarrow ZENO() \geq 1 \ \&\& \ ZENO() = Card(script) \\ C4.3 &:= R1 \ \&\& \ DD() \end{aligned}$$

C5.1 Représenter des objets ou séquences d’actions au moyen d’une représentation formelle : si une mission demande au joueur de combiner (DD) des blocs d’action (IEQ).

$$C5.1 := DD() \ \&\& \ IEQ(\{AllActions\}, 1)$$

C5.2 Traduire des objets ou séquences d’actions entre représentations formelles : si une mission cache la position du téléporteur de sortie.

$$C5.2 := Card(fog) \geq 1 \ || \ Card(hideExits) \geq 1$$

C6.1 Vérifier si une séquence d’actions atteint un objectif donné : si une mission contient un robot à programmer.

$$C6.1 := Card(robot) \geq 1$$

C6.2 Repérer des erreurs dans une séquence d’actions : aucune mécanique de jeu pour cette compétence.

$$\text{Aucune mécanique de jeu pour cette compétence.}$$

C6.3 Corriger une séquence d’actions pour atteindre un objectif donné : si une mission propose seulement des zones d’édition pré-construites boguées ($R1$) que le joueur doit corriger (DD).

$$\begin{aligned} R1 &\rightarrow ZEB() \geq 1 \ \&\& \ ZEB() = Card(script) \\ C6.3 &:= R1 \ \&\& \ DD() \end{aligned}$$

C6.4 Étendre ou modifier une séquence d’actions pour atteindre un nouvel objectif : si une mission demande au joueur de combiner (DD) des blocs d’action (IEQ) pour contrôler plusieurs robots avec une même zone d’édition ($R1$).

$$\begin{aligned} R1 &\rightarrow Card(robot \ inputLine \ sameValue \ robot \ inputLine) \geq 2 \\ C6.4 &:= DD() \ \&\& \ IEQ(\{AllActions\}, 1) \ \&\& \ R1 \end{aligned}$$

Framework de conception et d'implémentation de générateurs d'activités de jeu d'entraînement aux connaissances déclaratives

Framework for the design and implementation of game activity generators for declarative knowledge training

Bérénice LEMOINE¹; Pierre LAFORCADE¹; Sébastien GEORGE¹

¹Laboratoire d'Informatique de l'Université du Mans, Laval, France

Résumé. La génération d'activités adaptées est une technique peu abordée en Environnements Informatiques pour l'Apprentissage Humain. Cet article aborde l'aide à la conception de générateurs d'activités adaptées dans les jeux sérieux pour l'entraînement aux connaissances déclaratives. La proposition est un *framework* de conception de générateurs d'activités (i.e., architecture logicielle), orienté modèles et méta-modèles, extensible à différents domaines didactiques. Les générateurs produits permettent d'obtenir des activités d'entraînement adaptées et variées sous forme de niveaux de donjon pour des jeux de type *Roguelite*. L'article présente ce *framework* ainsi que son extension à deux domaines didactiques (i.e., table de multiplication et faits de judo) et son évaluation du point de vue ingénierie (i.e., tests automatisés, validation de modèles).

Mots-clés : génération, adaptation, jeux sérieux, conception, modélisation

Abstract. *Activity generation is a relatively underexplored technique in Technology Enhanced Learning. This article targets the design assistance of game activity generators for declarative knowledge training. The proposal is a model-driven framework for activity generator design (i.e., software architecture) that can be extended to various didactic domains. The produced generators provide varied and adapted training activities in the form of dungeon levels for Roguelite-oriented games. The article presents this framework, its extension to two didactic domains, and its engineering evaluation, encompassing automated testing and model validation.*

Keywords: *generation, adaptation, serious games, design, modelling*

1. INTRODUCTION

La mémorisation à court et long termes de connaissances déclaratives (e.g., lois, faits, règles) nécessite de la répétition (Kim *et al.*, 2013). Cependant, la répétition peut rapidement devenir ennuyeuse pour les apprenants (Smith, 1981). Toutefois, les jeux sérieux offrant des activités redondantes avec un challenge non adapté aux compétences/connaissances des apprenants-joueurs peuvent également conduire à un sentiment d'ennui (Streicher et Smeddinck, 2016) pouvant pousser à l'abandon des tâches, impactant alors l'apprentissage. En conséquence, pour limiter le sentiment d'ennui, les jeux sérieux visant l'entraînement de connaissances déclaratives doivent proposer des activités 1) variées et 2) adaptées aux apprenants-joueurs. La personnalisation manuelle d'activités pour chaque apprenant est une tâche chronophage et exigeante en termes d'efforts. De plus, concevoir des activités de jeu nécessite des compétences en *game design* (e.g., conception de situations de jeu) que les enseignants ne possèdent pas forcément.

D'autre part, la littérature en psychologie cognitive a montré que le processus de récupération de concepts ou de faits par le biais de tests augmente leur acquisition à long terme (Brame et Biel, 2015). Le concept de *Retrieval Practice* est une forme d'apprentissage par le test, consistant en des rappels répétés de ce qui a été appris (e.g., par l'utilisation de flashcards, de quiz) (Roediger et Pyc, 2012). Dans notre contexte, nous définissons l'entraînement comme une forme de *Retrieval Practice* consistant à poser, de manière répétée, différentes formes de questions sur des faits aux apprenants-joueurs.

La génération est une technique informatique (voir Figure 1) permettant la création automatique de contenu (e.g., niveau de jeu, histoire, dialogue) à partir d'un ensemble de données structurées et d'un ensemble de règles définies au travers d'algorithmes. Par conséquent, la génération d'un contenu décrivant une activité d'entraînement formalisée, à destination d'un interpréteur inclus dans le jeu sérieux d'entraînement, est une solution envisageable pour proposer des activités variées et adaptées. Cependant, le principe de génération n'est que très peu abordée dans le domaine des Environnements Informatiques pour l'Apprentissage Humain (EIAH) (Bezza *et al.*, 2013).

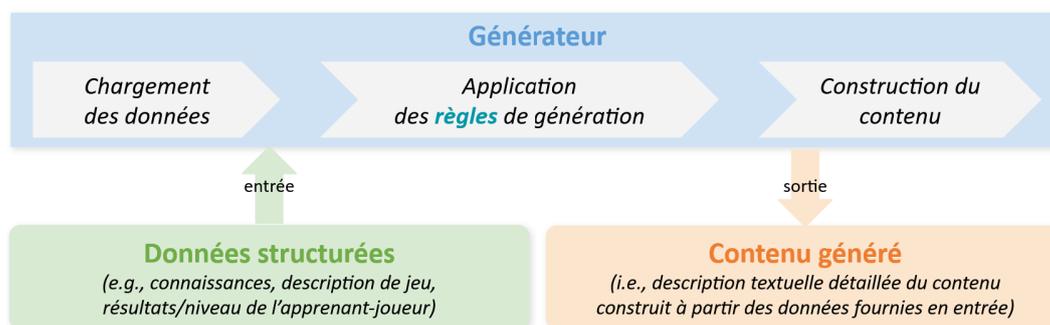


Figure 1 : Principe de génération de contenu

La structure d'une activité de jeu est entièrement dépendante du genre de jeu visé. Par exemple, une activité consistant à résoudre des énigmes et une activité consistant à explorer un monde ne se construisent pas de la même façon. Une activité d'énigme va être construite à partir d'un problème à résoudre (e.g., *Professeur Layton*) alors qu'une activité consistant à explorer va nécessiter la création d'un monde (zone d'exploration virtuelle) et d'un avatar (personnage) qui se déplacera et effectuera des actions. En conséquence, pour générer des activités de jeu d'entraînement, un genre de jeu doit être sélectionné. Les *Roguelites* sont souvent des jeux de type exploration de donjons dans lesquels les joueurs doivent parcourir des

niveaux générés, combattre des ennemis, collecter des objets et progresser. Ce genre de jeu a été analysé et il semble posséder les différentes caractéristiques nécessaires à de l'entraînement aux connaissances déclaratives : variété (génération procédurale avec de l'aléatoire), répétition (mécanique de mort permanente) et progression (rétention d'items) (Lemoine *et al.*, 2024b).

Nos travaux portent sur la recherche en ingénierie pour la conception et le développement de générateurs. Ces générateurs doivent produire des activités de jeu de type *Roguelite*, variées et adaptées pour l'entraînement aux connaissances déclaratives. En considérant les jeux d'entraînement comme des logiciels, les générateurs d'activités sont des composants logiciels du point de vue informatique. Nos travaux consistent en une recherche exploratoire visant à mieux caractériser ces générateurs (objets de recherche) et à proposer des modèles, des outils et des techniques pour faciliter leur conception (au sens informatique) et leur développement. Concevoir de tels générateurs est complexe et ne peut pas être réduit à un pur problème d'ingénierie informatique (Tchounikine *et al.*, 2009). En effet, de nombreux problèmes de spécifications et de mise en œuvre requièrent l'expertise de différents acteurs : les développeurs du jeu d'entraînement (choix de conception, exploitation des technologies), les experts du domaine didactique visé (faits à travailler, manière de les travailler, choix des adaptations par apprenant) et des experts en jeux vidéo (connaissances en jeu, *game design*...).

Dans ce but, nous proposons un *framework* (i.e., une infrastructure logicielle composée de modèles et d'outils) permettant d'assister la conception et l'implémentation de ces générateurs. Le travail présenté dans Lemoine *et al.* (2023) avait pour objectif de présenter le *framework* en se focalisant sur la dimension éducative uniquement. Cet article étend ce travail en y ajoutant les éléments en lien avec la dimension ludique. La proposition s'inscrit dans le cadre d'une approche d'Ingénierie Dirigée par les Modèles (IDM) (Kent, 2002). Le *framework* a permis de produire un générateur d'activités d'entraînement aux tables de multiplication qui est actuellement utilisé dans un prototype de jeu.

La Section 2 présente les travaux existants. La Section 3 définit notre contexte de recherche et positionne nos travaux. La Section 4 présente le *framework* dans son ensemble. La Section 5 présente une preuve de concept de l'extensibilité du *framework* au travers de son application à deux domaines didactiques. La Section 6 présente l'évaluation du *framework* en termes d'infrastructure logicielle. Enfin, la Section 7 conclut et présente les perspectives.

2. ÉTAT DE L'ART

La génération d'activités adaptées pour des jeux d'entraînement nécessite de s'intéresser à trois axes de recherche : la conception de jeux (plus particulièrement, la conception d'activités de jeu), l'adaptation de contenu, et la génération de contenu. Il est important de souligner que **la conception de générateurs d'activités de jeu ne nécessite pas de concevoir des jeux sérieux**. Les générateurs sont des composants logiciels indépendants, qui peuvent donc être évalués indépendamment d'un jeu. Cependant, créer des activités de jeux nécessite d'effectuer en amont des choix en termes de conception de jeu.

2.1. CONCEPTION DE JEUX ET DE JEUX SÉRIEUX

La conception de jeux et de jeux sérieux a fait l'objet de nombreuses méthodes, *frameworks* et approches (Amory, 2007 ; Carvalho *et al.*, 2015). Parmi ces travaux, on peut citer le *framework* MDA (*Mechanics, Dynamics, Aesthetics*) (Hunicke *et al.*, 2004) qui propose de décomposer les jeux en trois parties : la mécanique correspondant aux principaux com-

posants du jeu, la dynamique correspondant au comportement du jeu en cours d'exécution et l'esthétique correspondant aux réactions émotionnelles souhaitables pour le joueur. Ou encore, GOM (*Game Object Model*) version 1 et 2 (Amory, 2007), un *framework* orienté objets qui considère que le jeu sérieux est composé d'un ensemble de composants appelés *objets* qui sont décrits par des interfaces abstraites et concrètes. Les éléments éducatifs sont associés aux interfaces abstraites et les éléments de jeux sont associés aux interfaces concrètes. Cependant, la majorité des travaux existants sont orientés soit vers l'analyse de jeux existants, soit vers l'assistance à la conception générale (conception de haut niveau et spécification des besoins) du jeu à développer (Junior et Silva, 2021). À notre connaissance, aucun de ces travaux de conception de jeux et de jeux sérieux n'aborde la notion d'adaptation aux apprenants-joueurs, ni la notion de génération. Pourtant, de nombreux jeux commerciaux sont fondés sur un principe de génération (e.g., Hadès, *The Binding of Isaac*, *Enter the Gungeon*).

2.2. ADAPTATION ET MÉTHODES D'ADAPTATION EN EIAH

L'adaptation peut être mise en œuvre de différentes manières et peut viser une ou plusieurs cibles (e.g., les préférences de jeu, le contenu d'apprentissage, la difficulté). L'adaptation est souvent caractérisée par trois concepts : la source (à quoi adaptons-nous ?), la cible (qu'est-ce qui est adapté ?) et les « chemins » (comment est adaptée la cible à la source ?) (Vandewaetere *et al.*, 2011). Les formes d'adaptation qui nous intéressent principalement sont l'individualisation et la personnalisation de contenu. L'*individualisation* peut être définie 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 (Grant et Basye, 2014). D'autre part, la *personnalisation* peut être définie comme l'utilisation de modèles dans le but d'adapter les systèmes à chaque personne (Bakkes *et al.*, 2012 ; Ismail et Belkhouche, 2018)¹. De nombreux travaux proposent d'adapter la ludification de plateformes ou contenus éducatifs (Codish et Ravid, 2015 ; Monterrat *et al.*, 2017). Cependant, la ludification de contenus et la conception de jeux sérieux présentent différents objectifs. Ludifier consiste à ajouter une couche d'éléments de jeux à du contenu éducatif déjà existant et structuré. En revanche, concevoir un jeu sérieux requiert de penser conjointement la structuration du contenu éducatif et la structuration du jeu (Prensky, 2005).

Plusieurs travaux abordent la conception ou l'aide à la conception de contenus de jeux adaptés. Natkin *et al.* (2007) proposent un système de recommandation de quêtes en fonction d'un modèle utilisateur. Un ensemble de quêtes, sélectionné à partir du modèle utilisateur, est proposé au joueur. En fonction des choix du joueur et de ses traces d'utilisation, le modèle est mis-à-jour pour permettre de raffiner les propositions de quêtes par la suite. Marne *et al.* (2013) présentent un outil auteur permettant de soutenir les enseignants dans la conception de scénarios de jeux adaptés et non linéaires. En fonction de leurs réponses, les apprenants-joueurs sont confrontés à différents scénarios. En plus des activités d'apprentissage, les scénarios incluent des activités purement ludiques. Marty et Carron (2011) présentent une approche permettant de créer des modèles usager adaptés pour les jeux sérieux. Bontchev *et al.* (2021) proposent un framework incluant un modèle d'apprenant-joueur pour la personnalisation de jeux sérieux de type labyrinthe. Ce modèle est découpé en trois axes : les caractéristiques du joueur, les caractéristiques de l'apprenant et les caractéristiques de l'utilisateur.

Ces travaux ont tous un point commun : ils utilisent des modèles et des résultats des apprenants pour adapter. Cependant, les formes d'adaptation visées sont toutes différentes :

1. Ces deux termes seront traités comme des synonymes dans le cadre de cet article.

système de recommandation fondé sur un modèle utilisateur (Natkin *et al.*, 2007), système de sélection d'activités à partir des réponses utilisateurs et du scénario créé par l'enseignant (Marne *et al.*, 2013), modèles incluant des paramètres permettant l'adaptation aux apprenants (Bontchev *et al.*, 2021 ; Marty & Carron, 2011). Les approches étant différentes, les modèles sont dépendants des contextes et donc difficilement applicables à d'autres contextes. Cependant, ces approches peuvent aider ou orienter la définition d'autres modèles similaires dans d'autres contextes.

Quelques travaux s'intéressent à assister la conception de systèmes d'apprentissage adaptés. Par exemple, Roepke *et al.* (2021) proposent une architecture modulaire, basée sur des composants, pour mettre en œuvre des pipelines personnalisés pour les jeux d'apprentissage dans le cadre de la formation à la lutte contre l'hameçonnage. Un pipeline est un processus en trois étapes : collecte de données, génération de contenu, diffusion de contenu. Ismail et Belkhouche (2018) proposent une architecture réutilisable pour la conception de systèmes logiciels d'apprentissage personnalisé, décomposée en quatre unités : l'unité apprenant (qui conserve les données relatives à l'apprenant), l'unité des connaissances (qui conservent les ressources d'apprentissage), l'unité de personnalisation (qui fait correspondre le modèle de l'apprenant aux ressources d'apprentissage) et l'unité de présentation (qui représente l'environnement du logiciel). Ces travaux proposent des lignes directrices suffisamment générales pour être suivies pour la conception de tout système d'apprentissage adapté.

2.3. GÉNÉRATION EN EIAH ET DANS LES JEUX

Bien que peu traitée en EIAH, la problématique de la génération, consistant à construire automatiquement du contenu, a été abordée sous trois angles principaux : génération de contenu non adapté, génération de contenu adapté à l'apprenant et génération de contenu adapté au joueur. Diwan *et al.* (2019) proposent un modèle pour générer des parcours pédagogiques à partir de ressources open-source. Holohan *et al.* (2006) ont défini une ontologie permettant de décrire les bases de données relationnelles. Cette ontologie est utilisée pour générer automatiquement des exercices en ligne pour l'apprentissage de connaissances procédurales (i.e., les bases de données relationnelles). Carpentier et Lourdeaux (2014) proposent une approche pour générer dynamiquement des scénarios adaptés aux capacités des apprenants et aux besoins pédagogiques dans des environnements virtuels. Leur approche s'inscrit dans un cadre fondé sur trois modèles : le modèle du domaine (la description statique du monde, de ses éléments et de leurs relations), le modèle de l'activité (la structure hiérarchique de l'activité observée), le modèle de causalité (l'expression des chaînes causales pertinentes se produisant dans l'environnement). Sehaba et Hussaan (2013) proposent une architecture générale pour la génération de scénarios de jeu adaptés à l'apprenant (c'est-à-dire à ses compétences, aptitudes et besoins). Cette architecture repose sur plusieurs modèles : le modèle du domaine qui modélise les concepts du domaine et leurs relations ; le modèle de l'apprenant qui modélise les informations personnelles, la motivation, les compétences et les interactions des apprenants ; le modèle de présentation qui décrit les structures des scénarios ; le modèle du jeu sérieux qui associe les ressources du jeu à celles de la pédagogie. Les connaissances en matière d'adaptation sont représentées sous la forme d'un système basé sur des règles. Laforcade et Laghouaouta (2018) proposent une approche IDM, inspirée des travaux de Sehaba et Hussaan (2013), pour permettre la spécification des générateurs de séquences d'activités (scénarios de jeu) adaptées aux besoins individuels de l'apprenant. L'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). Callies *et al.* (2015) proposent une architecture adaptative consistant à générer des plans pédagogiques adaptés ainsi qu'à

adapter les comportements des personnages non-joueurs en fonction des actions des joueurs dans des jeux de type simulation. Ce travail est basé sur un modèle de joueur composé de ses connaissances du jeu et du domaine, ainsi que sur un module d'adaptation.

Dans le contexte des jeux, certains travaux abordent la génération de contenus adaptés, comme Dormans et Bakkes (2011) 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 (modèle de joueur) afin de personnaliser l'expérience de jeu. Sina *et al.* (2014) proposent *ScenarioGen* une méthode pour générer du contenu textuel sur les activités quotidiennes. Leur méthode consiste à concevoir un nouveau scénario à partir d'un profil utilisateur et d'un scénario existant. Ces travaux ont tous un point commun : ils utilisent des données structurées (e.g., modèles, ontologies) à partir desquelles du contenu est généré.

2.4. CONSTATS

Le point essentiel de ces recherches réside dans le fait que l'adaptation et la génération de contenu sont majoritairement fondées sur l'utilisation de données structurées telles que des modèles, des ontologies, etc. En ce qui concerne l'adaptation du contenu éducatif, avec ou sans génération, les principales données structurées utilisées sont :

- des informations sur le domaine qui décrivent les connaissances et le contenu d'apprentissage ;
- des informations sur l'apprenant comprenant ses données personnelles, sa progression, ses résultats, etc. ;
- des informations sur la structure des activités à générer ; et des informations sur les règles d'adaptation.

D'autre part, pour ce qui est de l'adaptation du contenu des jeux, avec ou sans génération, les principales données structurées utilisées sont :

- des informations sur le jeu, c'est-à-dire la description statique des éléments du jeu ;
- des informations sur la structure de l'activité de jeu à générer ;
- et des informations sur le joueur incluant ses préférences, sa progression de jeu, etc.

Lorsque ces travaux se penchent sur la génération de contenu de jeu adapté, ils introduisent généralement des informations supplémentaires concernant les relations entre les éléments de jeu et les éléments éducatifs. Cependant, ces relations sont souvent intégrées au sein de l'algorithme de génération sans être spécifiées à travers des modèles de manière explicite. En outre, les modèles ou méthodes de représentation des données proposés sont généralement spécifiques à un contexte ou à un domaine didactique particulier, ce qui les rend difficilement réutilisables pour d'autres domaines.

Un second constat important est que l'adaptation est souvent abordée soit du point de vue éducatif, soit du point de vue ludique, mais rarement dans les deux perspectives simultanément. L'ajustement des activités en fonction des apprenants est reconnu comme un moyen d'améliorer l'apprentissage. De plus, l'adaptation aux joueurs contribue à rendre les tâches plus engageantes et motivantes. Il semble donc pertinent d'envisager l'adaptation en prenant en considération à la fois la dimension éducative et ludique.

Par conséquent, la création d'activités de jeu pour l'entraînement nécessite la création de modèles pour spécifier l'ensemble des données nécessaires comme le domaine didactique visé, la structure de l'activité, les résultats et la progression des apprenants-joueurs ou encore les éléments de jeu. Afin de favoriser un apprentissage plus efficace et un engagement accru des apprenants-joueurs, la génération doit offrir des activités adaptées sur les dimensions éducative et ludique.

3. CONTEXTE DE RECHERCHE

3.1. PROJET ADAPTABLES

AdapTABLES est un projet de recherche qui s'intéresse à l'acquisition longue durée des tables de multiplication. Ce projet vise à concevoir et développer un jeu sérieux dédié à l'entraînement aux tables de multiplication. L'objectif principal du projet est de stabiliser les connaissances (Dias, 2018), c'est-à-dire que les tables sont considérées comme déjà expérimentées et comprises.

De nombreux jeux sérieux proposent de travailler les tables de multiplication en ligne. Cependant, ces jeux présentent en général des questions auxquelles les joueurs doivent répondre, accompagnées de mécaniques de jeu (e.g., récompenses, scores, pression du temps). Les choix en termes de pédagogie se limitent souvent à la sélection des tables à travailler ou au niveau de difficulté (facile, moyen, difficile). Ces niveaux de difficultés ont majoritairement un impact sur le temps de réponse autorisé ou sur les tables travaillées. Cependant, la méthode d'attribution des niveaux de difficulté aux tables n'est pas explicitée. Dans le cadre du projet, une étude exploratoire (Laforcade *et al.*, 2022) a été conduite à l'aide d'experts en mathématiques (i.e., enseignants des cycles 2-3 et didacticien). Le principal objectif de cette étude était de permettre la spécification des besoins d'adaptation du point de vue des tables de multiplications (organisation d'un entraînement, tâches d'entraînement, éléments à faire varier, etc.).

3.2. PROBLÉMATIQUE ET POSITIONNEMENT

À partir des constats relevés précédemment, notre problématique est la suivante : **comment faciliter la conception de générateurs d'activités de jeu adaptées et variées, destinées à l'entraînement aux connaissances déclaratives?** Plusieurs questions de recherche découlent de cette question générale :

1. Comment proposer une approche suffisamment générique pour considérer les connaissances déclaratives indépendamment d'un domaine didactique spécifique ?
2. Qu'est-ce qu'une activité d'entraînement de jeu adaptée et variée ? De quels éléments éducatifs et de quels éléments de jeu sont composées ces activités ? Comment associer les éléments de jeu et les éléments éducatifs de manière cohérente ?
3. Comment structurer ces éléments et leurs relations pour guider la génération d'activités cohérentes ?
4. Comment spécifier ces informations informatiquement pour développer des générateurs d'activités ?

L'intérêt de s'intéresser aux connaissances déclaratives réside dans la possibilité de proposer des modèles pouvant être réutilisés au-delà d'un domaine didactique spécifique. Comme précédemment mentionné, l'acquisition de connaissances déclaratives implique une pratique répétée (Kim *et al.*, 2013). Dans notre contexte, l'entraînement consiste en la répétition de diverses formes de questions sur des faits, qui sont posées aux apprenants-joueurs de manière répétée. Or, la structure d'une activité de jeu étant dépendante du genre de jeu visé, un genre doit être sélectionné pour permettre la génération d'activités de jeu d'entraînement. En conséquence, 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 (i.e., conditions nécessaires à l'entraînement). Le *Roguelite* répond à ces besoins (Lemoine *et al.*, 2024b). Ce genre 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*, ...) facilitant la progression ludique dans la prochaine partie. Par conséquent, une activité de jeu d'entraînement est un *donjon*, c'est-à-dire un ensemble de salles interconnectées dans lesquelles l'avatar se déplace et où l'entraînement a lieu.

Pour réduire la sensation de répétition causée par l'entraînement, les activités générés doivent être variées et adaptées. Étant donné que l'adaptation aux deux dimensions (éducative et ludique) est rarement abordée conjointement, notre objectif est d'individualiser les activités pour les joueurs et les apprenants. Pour être plus précis, l'**adaptation** vise à prendre en considération **trois perspectives** distinctes : celle de l'**enseignant** (stratégies d'entraînement, choix pédagogiques et didactiques), celle de l'**apprenant** (son niveau de connaissance et sa progression), et celle du **joueur** (ses préférences de jeu). L'adaptation envisagée, selon Plass et Pawar (2020), porte sur deux variables cognitives (le niveau ou la progression de l'apprenant et la stratégie de l'enseignant) ainsi qu'une variable motivationnelle (l'intérêt du joueur à travers ses préférences). Ainsi, l'adaptation éducative consistera à prendre en compte divers aspects tels que les types de faits rencontrés, la manière de les questionner, leur nombre, leur ordre en fonction du niveau de l'apprenant, de ses résultats antérieurs et de l'entraînement établi par l'enseignant. D'autre part, les *Roguelites* possèdent souvent une mécanique d'achat ou d'activation permettant aux joueurs d'effectuer des choix pouvant ou non influencer la génération des niveaux de jeu. Ainsi, l'adaptation des préférences consistera à se focaliser sur la possibilité d'activer ou de désactiver des éléments de jeu, notamment des équipements débloquent différents *gameplays* qui permettront de répondre aux faits questionnés, en orientant ou en déplaçant des objets, par exemple. Ce choix de conception permet aux joueurs de désactiver des *gameplays* qui leur déplaisent.

En plus de l'adaptation, la variété des activités est nécessaire pour réduire le sentiment d'ennui produit par la répétition. Les *Roguelites* sont fondés sur un mécanisme de génération procédurale avec de l'aléatoire rendant chaque niveau de jeu différent en termes de contenus (structure des donjons, éléments et leur position dans les salles, etc.). Dans notre contexte, notre approche consiste à modéliser des éléments (éducatifs et de jeux) avec une certaine variabilité. Par exemple, les faits sont modélisés pour que les mauvais choix soient définis à chaque génération d'un donjon. Ou encore, les *gameplays* sont définis à l'aide de **capacités** pour permettre de faire varier les éléments de jeux sélectionnés à chaque génération d'un donjon. L'objectif est de permettre à l'algorithme de génération de choisir de façon pseudo-aléatoire, pour conserver la cohérence des activités, les éléments d'entraînement et de jeux d'un donjon.

Un générateur d'activité de jeu de type *Roguelite* pour l'entraînement aux connaissances déclaratives est un composant logiciel (élément constitutif d'un logiciel destiné à être incorporé en tant que pièce détachée) dont l'algorithme permet de construire des activités variées à partir de trois types d'informations fournies en entrée : des informations sur l'entraînement, des informations sur le jeu et des informations sur l'apprenant-joueur concerné. Ces composants logiciels produisent en sortie des descriptions détaillées d'activités (niveaux de donjon) adaptées aux apprenants-joueurs.

Pour aborder notre problématique, nous proposons un *framework* de conception et d'implémentation de générateurs d'activités adaptées et variées pour l'entraînement aux connaissances déclaratives à travers des jeux de type *Roguelite*. L'originalité de cette proposition et de ce positionnement est d'aborder l'adaptation en prenant en compte simultanément les dimensions de jeu et d'entraînement (apprentissage). De plus, le *framework* est un outil (infrastructure logicielle) disposant d'un mécanisme d'extension permettant de prendre en compte de nombreux domaines didactiques. À notre connaissance, aucune approche permet-

tant de guider la conception de générateurs au niveau algorithmique n'existe. Ce *framework* est composé d'un **cadre conceptuel** (Lemoine & Laforcade, 2023a) et d'une **infrastructure logicielle extensible** (partiellement présentée (Lemoine *et al.*, 2023)) à des domaines didactiques spécifiques. Cette infrastructure logicielle capture l'ensemble des éléments communs pour tout domaine didactique. Tandis que le mécanisme d'extension guide l'ajout des éléments spécifiques à un domaine didactique visé. L'avantage d'une approche extensible est de faciliter l'implémentation de générateurs d'activités, par des ingénieurs ou développeurs, en limitant le développement nécessaire aux informations reliées aux connaissances déclaratives du domaine didactique. Ce framework propose un algorithme, des modèles, et méta-modèles déjà existants à étendre et qui seront donc réutilisés pour spécifier un générateur d'activités.

Notre proposition s'inscrit dans le contexte d'une approche fondée sur l'Ingénierie Dirigée par les Modèles (Kent, 2002). L'IDM est fondée sur la notion de modèle (abstraction d'un système selon un point de vue) et repose sur quatre principes : la capitalisation (les modèles doivent être réutilisables), l'abstraction (les modèles doivent être indépendants des technologies), la modélisation (les modèles doivent adopter une vision productive, c'est-à-dire permettre la génération de code final du logiciel), et la séparation des préoccupations (Jézéquel *et al.*, 2012). L'IDM est un champ de recherche vaste abordant la spécification, l'exécution, la transformation et la composition de modèles. La transformation de modèles est une opération centrale en IDM qui permet la génération automatique de modèles à partir de modèles sources. Pour permettre ces transformations, les modèles doivent être conformes à des méta-modèles (modèle permettant de décrire des modèles) qui définissent la structure et les règles qu'ils doivent respecter. Les principes de l'IDM, ainsi que les différents outils permettant de soutenir son utilisation, en font une approche très intéressante dans notre contexte. La contribution présentée s'inscrit dans le cadre de la recherche en ingénierie des systèmes EIAH (Tchounikine *et al.*, 2009) contribuant à l'exploration et à l'orientation des solutions pour la génération d'activités adaptées.

4. FRAMEWORK DE CONCEPTION ET D'IMPLÉMENTATION DE GÉNÉRATEURS

Cette section présente la contribution : un framework de conception et d'implémentation de générateurs d'activités de jeu d'entraînement aux connaissances déclaratives. La sous-section 4.1 présente une vue d'ensemble de la proposition. La sous-section 4.2 présente le cadre conceptuel du *framework* défini au cours de cette recherche exploratoire. Enfin, la sous-section 4.3 présente l'implémentation informatique du *framework* (modèles interprétables, algorithme de génération, règles d'extension).

4.1. PRÉSENTATION GÉNÉRALE ET PROPRIÉTÉS

Le *framework* proposé est une infrastructure conceptuelle et logicielle composée d'un ensemble de modèles et d'outils pour formaliser et pour guider l'implémentation de générateurs d'activités variées et adaptées. Les générateurs produits sont des éléments logiciels pouvant être considérés comme des composants d'un jeu d'entraînement aux connaissances déclaratives de type *Roguelite*. Ces générateurs permettent de produire une nouvelle activité d'entraînement à chaque demande, c'est-à-dire une description textuelle détaillée d'un niveau de donjon, pour un apprenant-joueur donné. Les descriptions doivent ensuite être interprétées par un *game player* pour proposer un niveau de jeu jouable à l'apprenant-joueur.

Le *framework* est fondé sur une approche conceptuelle structurant l'ensemble des informations nécessaires à la génération (voir Section 4.2) et implémenté au travers d'une approche d'Ingénierie Dirigée par les Modèles (voir Section 4.3). Ce *framework* se décompose en deux parties : un ensemble de composants génériques et des règles permettant l'extension à des domaines didactiques spécifiques. Étant dans le cadre d'une approche IDM, les composants génériques regroupent un ensemble de modèles et de méta-modèles ainsi qu'un algorithme de génération d'activités adaptées et variées utilisant ces modèles. Comme nous l'expliquerons ultérieurement, certains éléments ne peuvent pas être traités ou générés indépendamment du domaine. Les règles d'extension permettent de construire les modèles, les méta-modèles, et les composants de code spécifiques au domaine et requis par le générateur. La Figure 2 présente les différents composants du *framework*.

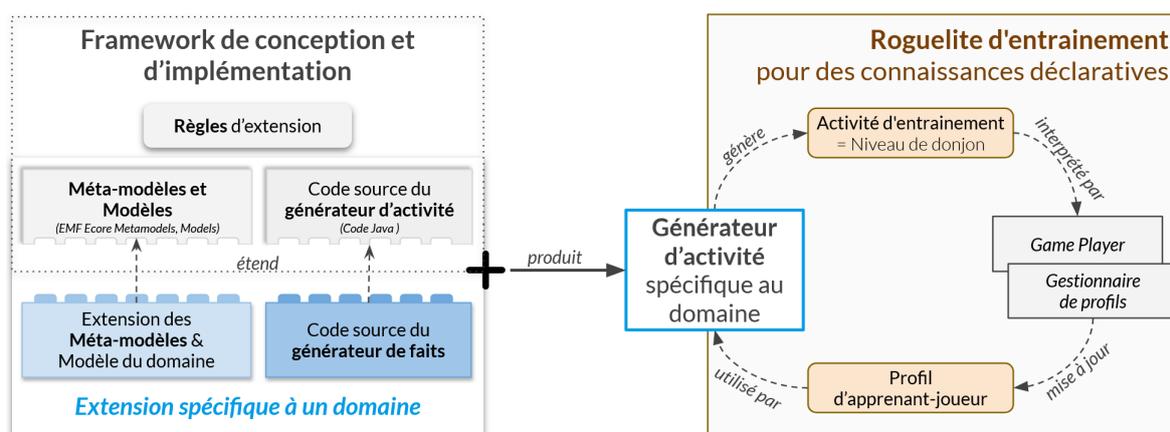


Figure 2 : Positionnement général du *framework* de conception

Dans un contexte IDM, l'algorithme de génération peut être perçu comme réalisant une transformation de modèle. Ainsi, les générateurs produits requièrent en entrée, un modèle conforme à chaque méta-modèle d'entrée du *framework* (données concrètes : connaissances à travailler, éléments de jeu disponible, résultat et progression de l'apprenant-joueur, etc.) et produisent en sortie un modèle également conforme à un méta-modèle du *framework*. Ces générateurs sont composés du code source correspondant aux méta-modèles du *framework*, leur permettant de lire et d'utiliser les données contenues dans les modèles. Ils possèdent également l'algorithme de génération d'activité générique (indépendant du domaine didactique) ainsi que les modèles et l'algorithme de génération de faits conçu lors de l'extension du *framework* en fonction du domaine didactique. La Figure 3 représente l'ensemble des composants d'un générateur dans notre contexte.

Comme évoqué dans la Section 3.2, les activités générées doivent être variées et adaptées en fonction de trois points de vue distincts : celui de l'enseignant, en tenant compte de sa vision de l'entraînement pour chaque apprenant, celui de l'apprenant, en prenant en considération son niveau et sa progression dans l'entraînement, et enfin, celui du joueur, en intégrant ses préférences de jeu. De plus, le *framework* doit viser l'entraînement aux connaissances déclaratives non spécifiques à un domaine didactique. En conséquence, le *framework* et les générateurs produits par ce dernier doivent respecter certaines propriétés. Le *framework* doit posséder les deux propriétés suivantes :

- **FP1** : possibilité d'exprimer différents domaines didactiques ;
- **FP2** : possibilité d'exprimer la vision des enseignants sur l'entraînement des apprenants individuellement.

En ce qui concerne les générateurs, il est nécessaire de garantir le respect de trois propriétés :

- **GP1** : les activités générées doivent être adaptées au niveau et résultats de l'apprenant dans son parcours d'entraînement ;
- **GP2** : les activités générées doivent être adaptées aux préférences de jeu du joueur ;
- **GP3** : les activités générées doivent être variées sur le plan éducatif et de jeu.

Ce *framework* a été élaboré dans le cadre d'une recherche exploratoire centrée sur la conception itérative d'un cas d'étude initial, avec l'implication d'un groupe d'utilisateurs 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 deux autres domaines, c'est-à-dire les techniques et gestes d'arbitrages de judo et les repères d'histoire-géographie du Diplôme National du Brevet.

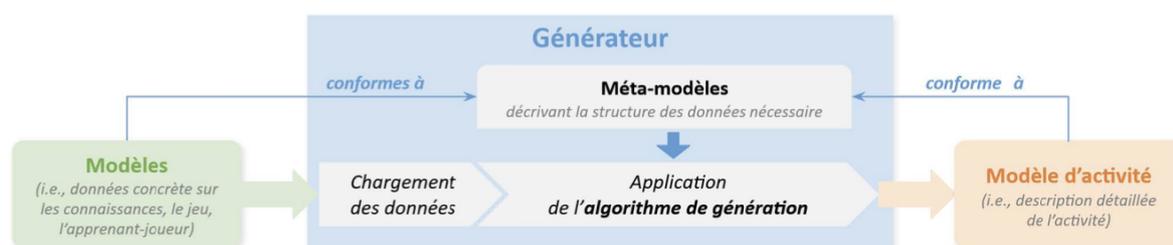


Figure 3 : Composants à gros grain d'un générateur d'activité d'entraînement de jeu

4.2. CADRE CONCEPTUEL : DES MODÈLES POUR LA GÉNÉRATION

Comme précédemment mentionné, la génération requiert différentes informations structurées pour générer des activités variées et adaptées. Nous avons proposé un ensemble de modèles conceptuels nécessaires à la génération d'activités de jeu de type *Roguelite* pour l'entraînement aux connaissances déclaratives (Lemoine et Laforcade, 2023a; 2023b). Ces modèles ont été définis pour être génériques (toutes connaissances déclaratives indépendamment d'un domaine spécifique). En conséquence, certains modèles contiennent des points d'extension, c'est-à-dire une portion de modèle à étendre en fonction du domaine visé. Cette section présente ces différents modèles conceptuels.

4.2.1. Modèle du domaine : parcours d'entraînement et connaissances

Dans le contexte du projet, une étude exploratoire a été menée avec des experts en mathématiques, partiellement présentée (Laforcade *et al.*, 2022). Cette étude avait deux principaux objectifs : 1) définir une structure de parcours d'entraînement, 2) spécifier les adaptations à prendre en compte pour l'entraînement aux tables de multiplication. À partir des résultats obtenus avec les experts en mathématiques, un travail d'abstraction nous a permis de définir une structure d'entraînement (voir Figure 4), indépendante des mathématiques, appelée parcours d'entraînement (Lemoine *et al.*, 2024b). Cette structuration vise à **explicitement la vision de l'enseignant sur la progression de l'entraînement** (adaptation), en termes de connaissances à travailler et de paramètres à faire varier, **pour un apprenant ou un groupe d'apprenants**. Un parcours d'entraînement consiste en un ensemble d'objectifs (e.g., s'entraîner sur la table de 2) ordonnés par des relations de pré-requis. Chaque objectif vise un ensemble de faits à travailler (connaissances visées) et est décomposé en niveaux progressifs, eux-mêmes décomposés en tâches d'entraînement.

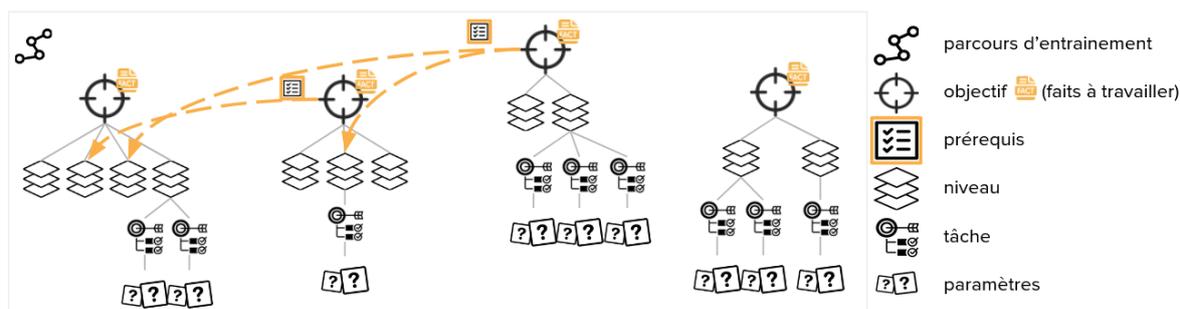


Figure 4 : Illustration de la structure des parcours d'entraînement

Dans le cadre des tâches pour l'entraînement aux tables de multiplication, cinq tâches ont été identifiées :

- Complétion 1 : compléter un fait incomplet ayant un élément manquant (e.g., $3 \times ? = 15$, $15 = ? \times 5$);
- Complétion 2 : compléter un fait incomplet ayant deux éléments manquants (e.g., $3 \times ? = ?$ avec un ensemble de choix donnés [3, 6, 5, 15]);
- Reconstruction : 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 : identifier l'exactitude ou l'inexactitude d'un ou de plusieurs faits (e.g., $3 \times 5 = 15$, vrai ou faux ?);
- Identification d'Appartenance : 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 ?).

Notons que les trois premières tâches (Complétion 1, Complétion 2, Reconstruction) sont trois formes de complétion de faits ayant des éléments manquants.

D'autre part, nous avons échangé avec des enseignants d'histoire-géographie² pour 1) discuter de la structure d'entraînement et 2) définir des tâches d'entraînement pour les repères d'histoire-géographie du brevet. Les enseignants avec lesquels nous avons discuté ont considéré la structure d'entraînement appropriée. Les tâches d'entraînement actuellement définies sont semblables à celles des mathématiques : association (une forme de complétion de plusieurs faits), vérification de la validité d'un fait (une forme de tâche d'identification), nommer et localiser sur une carte (une forme d'identification d'appartenance). Cependant, une tâche est différente, la tâche d'ordonnancement chronologique.

En fonction du domaine didactique, une tâche possède différents paramètres permettant de construire les faits à questionner. Pour illustrer, dans une tâche de complétion en mathématiques, l'élément manquant sera un entier (opérande, table ou résultat), tandis qu'en histoire, il s'agira plutôt d'une chaîne de caractères (événement, date ou période). Ainsi, pour permettre l'obtention d'un algorithme de génération d'activités générique et comme trois des tâches sont présentes en mathématiques et en histoire-géographie, nous avons défini **quatre types de tâches** génériques :

- **Complétion** (compléter des faits ayant des éléments manquants);
- **Identification** (identifier si des faits sont vrais ou faux);
- **Identification d'Appartenance** (identifier si des faits partagent ou non une propriété commune);
- **Ordonnancement** (ordonner des faits en fonction d'une heuristique donnée).

Ces tâches n'ont pas pour but d'être exhaustives, mais de couvrir au moins les domaines

2. Ce second domaine est présenté ici seulement pour illustrer d'autres faits et tâches spécifiques; dans la section dédiée à l'évaluation nous nous appuyerons sur un troisième domaine.

d'application visés.

Ainsi, un premier modèle à fournir au générateur est le **modèle du domaine** (*Learning-Domain*) qui comprend le parcours d'entraînement (*LearningPath*) et les connaissances à travailler. La Figure 5 présente une version simplifiée du modèle du domaine. Les faits à travailler sont regroupés en ensemble de faits (e.g., tables de multiplication) et peuvent être graphiques, c'est-à-dire avoir une représentation graphique, être associé à une image ou une carte (e.g., régions de France). Ces faits (*Fact*) sont des faits bruts tels que $3 \times 5 = 15$, puisqu'ils sont différents en fonction des domaines didactiques, la notion de fait est abstraite. De même, les niveaux (*Level*) sont également abstraits puisqu'ils peuvent posséder des paramètres spécifiques au domaine didactique (e.g., construction des tables de multiplication, 1×1 , $1 \times 2 \dots$ ou 1×1 , $2 \times 1 \dots$). Enfin, les tâches (*Task*) sont abstraites puisqu'elles possèdent des paramètres spécifiques au domaine didactique tel que le type d'élément recherché dans une tâche de complétion.

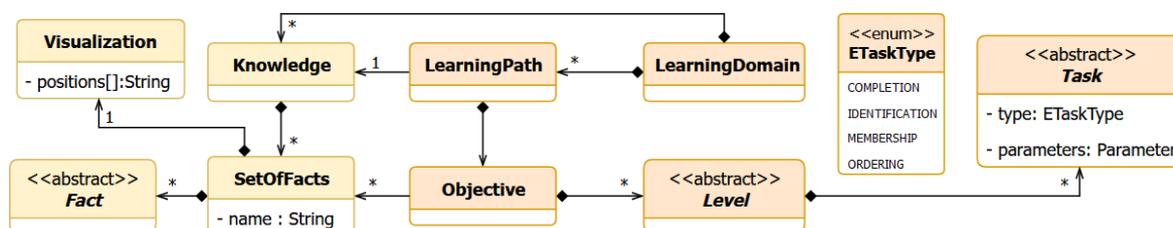


Figure 5 : Modèle conceptuel simplifié du modèle de domaine (Lemoine & Laforcade, 2023a)

Si certains paramètres des tâches sont spécifiques au domaine, d'autres sont communs à toutes les tâches ou à un type de tâche donné. Toutes les tâches décrivent le nombre de faits à questionner en même temps, le taux d'apparition de la tâche (favorisation d'une tâche plutôt qu'une autre par l'enseignant), le temps de réponse maximal attendu par question, le nombre de réponses attendues par fait, le critère d'acquisition des faits (combien de fois un fait doit être réussi consécutivement pour être considéré comme acquis ?). De plus, les tâches décrivent si ce sont des tâches graphiques (avec carte), si elles sont à validation automatique ou à validation manuelle de la réponse par l'apprenant (permet une correction et une auto-réflexion sur la réponse donnée, mais fait perdre du temps dans le cadre d'un jeu). Pour les tâches d'Identification d'Appartenance, un paramètre commun est un booléen indiquant si les éléments recherchés sont ceux qui partagent la propriété ou ceux qui ne la partagent pas.

4.2.2. Modèle du jeu

D'après Prensky (2005), la principale raison de l'échec des jeux d'apprentissage réside dans leur « manque » de *gameplay* (gameplays trop simples ou minimalistes). Dans un jeu, le *gameplay* est représenté par l'ensemble des éléments avec lesquels les joueurs interagissent (actions que l'avatar peut effectuer et qui ont un impact sur la progression) ou qui leur fournissent des informations. En conséquence, nous avons tenté de proposer une variété de *gameplays*.

Ainsi, en parallèle des échanges avec des experts des domaines, nous avons échangé avec des concepteurs de jeux vidéos pour concevoir des *gameplays* permettant de répondre à des questions sur des connaissances déclaratives (en s'inspirant des tâches définies dans les deux domaines). À partir des échanges, plusieurs maquettes de *gameplays* ont été définies. Suite à la conception des maquettes, nous avons constaté que certains *gameplays* semblaient appartenir à une même catégorie. Par exemple, casser un pot possédant la bonne réponse ou

ouvrir un coffre possédant la bonne réponse sont deux manières de sélectionner une réponse. Ce constat est cohérent avec la classification de jeux de Djaouti *et al.* (2008) qui consiste à décrire les jeux à l'aide de briques de *gameplay* (des catégories d'actions pouvant être réalisées dans les jeux). En conséquence, nous avons défini cinq catégories de *gameplay* dans le contexte des *Roguelites* d'entraînement aux connaissances déclaratives (ces catégories ne prétendent pas être exhaustives) : *SELECT* (sélectionner les objets avec les bonnes réponses), *MOVE* (déplacer les bons objets vers les zones attendues), *ORIENT* (orienter les objets vers les bonnes réponses), *POSITION* (placer l'avatar sur la bonne réponse) et *DIRECT RESPONSE* (saisir la bonne réponse) (Lemoine *et al.*, 2024a).

Ces catégories représentent différentes manières de répondre aux questions posées sur les faits. D'autres *gameplays* présentant des éléments présents dans les *Roguelites* comme les pièges à éviter, des ennemis, ou des objets à casser pour obtenir des pièces doivent également être présents pour conserver l'esprit du *Roguelite*. Cependant, si la définition statique des *gameplays* en termes d'éléments de jeu spécifiques permet un certain niveau de variété, elle impose deux contraintes : 1) elle est coûteuse en temps, c'est-à-dire que les *gameplays* doivent être décrits un par un en fonction des éléments de jeu disponibles et 2) elle est statique, c'est-à-dire que l'ajout d'un élément de jeu signifie qu'il faut spécifier de nouveaux *gameplays* pour cet élément. Ainsi, pour augmenter la variabilité des *gameplays* proposés, nous proposons de définir les *gameplays* et les éléments de jeux au travers de **capacités**. Les capacités définissent le comportement des éléments (la manière dont l'avatar du joueur peut interagir avec eux). Par exemple, un bloc peut être poussé (*pushable*), un pot peut être déplacé (*movable*), un pont peut être traversé (*crossable*), etc. Ainsi, plusieurs types d'éléments peuvent être définis avec une même capacité, par exemple, un cube et un pot peuvent être déplacés (*movable*). Cette modélisation permet de définir un *gameplay* d'éléments déplaçables mais de générer des *gameplays* jouables (description réelle du *gameplay* qui va être joué) avec différents types d'éléments. Ainsi, les *gameplays* sont décrits par des composants (*AComponent*), pouvant être simples (e.g., pot à casser) ou composites (e.g., bloc à pousser avec son détecteur associé), faisant référence à une capacité attendue. Ces composants décrivent également le type d'information attendu : proposition(s), question, propositions et question (e.g., structure à déclarer pour chaque question sur un fait), etc. Ces informations permettront de correctement instancier un *gameplay* (création des objets de jeu) en fonction des faits qui sont questionnés.

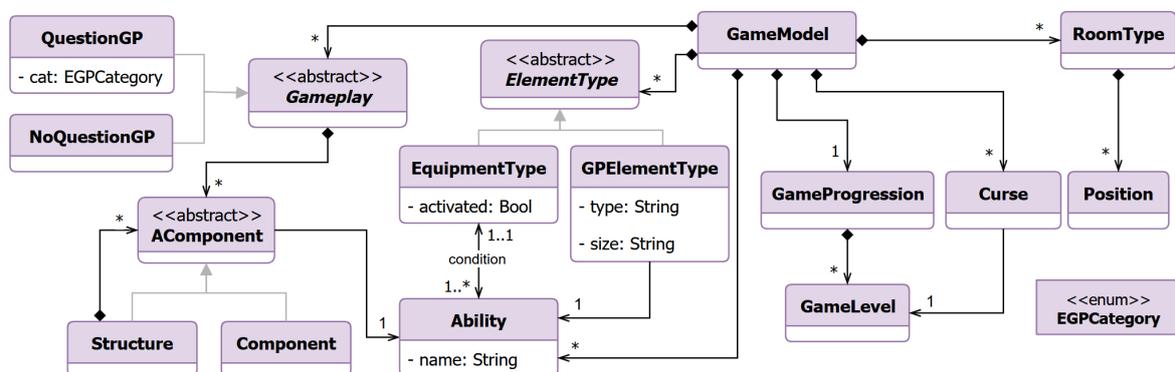


Figure 6 : Modèle conceptuel simplifié du jeu (Lemoine & Laforcade, 2023a, 2023b)

Comme précédemment précisé, les *Roguelites* sont fondés sur une mécanique d'achat ou d'activation. Dans notre contexte, l'achat d'équipement permet de déverrouiller des capacités et donc de débloquent de nouveaux *gameplays*. Ces éléments doivent donc être modélisés

pour permettre à l'algorithme de génération de ne choisir que des *gameplays* déverrouillés. D'autre part, dans le but de faire varier la forme des salles des donjons, nous avons fait le choix de modéliser des types de salles (*RoomType*) qui décrivent les positions possibles pour des éléments et les accès qu'ils autorisent. Enfin, un jeu possède une progression qui doit être spécifiée également. Comme pour la plupart des *Roguelites*, la progression est fondée sur une approche en termes de niveaux de difficultés. Un niveau de jeu influe sur la taille du donjon (le nombre de salles avec et sans question) et les pièges possibles (difficulté et nombre). De plus, de nombreux *Roguelite* propose une mécanique de malédiction qui apparaissent lorsque le joueur a atteint un certain niveau de difficulté. Par exemple, à partir du niveau 8 le niveau est dans le noir. Nous avons également fait le choix de conserver cette mécanique de malédiction. Cependant, certaines malédiction pourraient poser un problème du point de vue de l'apprentissage. Par exemple, une malédiction imposant de terminer un niveau en un temps donné pourrait aller à l'encontre de l'objectif pédagogique de l'enseignant. L'ajout de malédiction doit donc faire l'objet de réflexions avec les enseignants.

Ainsi, un second modèle à fournir au générateur est le modèle du jeu qui décrit l'ensemble des éléments de jeu disponibles et la progression à suivre du jeu. La Figure 6 présente une version simplifiée du modèle conceptuel du jeu.

4.2.3. Modèle de l'activité

Une activité d'entraînement aux connaissances déclaratives pour des *Roguelites* est un donjon, c'est-à-dire un ensemble de salles interconnectées. Un donjon a une entrée, une sortie et chaque salle a des accès vers d'autres salles (ses voisines). Une salle contient un *gameplay* de question (e.g., compléter le fait en déplaçant le bon pot) ou un *gameplay* de piège (e.g., éviter les pics pour passer à la salle suivante). En fonction des *gameplays* les salles vont posséder différents éléments positionnés. Un élément positionné est un élément de jeu qui possède la capacité spécifiée dans la description du *gameplay* sélectionné, et qui a été positionné dans le type de salle sélectionné. Ces éléments vont avoir différents paramètres. Lorsqu'ils représentent une partie d'une question sur un fait (e.g., proposition, question), les éléments pourront avoir une ou des valeurs à afficher, une information sur le type d'affichage (texte ou image), un paramètre de vérification (e.g., booléen décrivant si la proposition est correcte ou incorrecte), etc.

De plus, une salle est composée des faits questionnés qui sont interrogés dans cette salle. Les faits questionnés sont des questions à propos de faits. Comme précédemment mentionné, un fait brut a une structure différente en fonction du domaine didactique visé. Prenons l'exemple de deux tâches T1 et T2. T1 consiste à choisir parmi un ensemble de propositions la réponse correspondant au résultat de la multiplication pour chaque fait. À partir des paramètres de T1, les faits interrogés sont construits pour donner, par exemple, une question telle que $2 \times 6 = ?$ avec un ensemble de propositions tel que $\{8, 12, 14\}$. T2 consiste à choisir dans un ensemble de propositions celles qui sont des résultats possibles d'une table donnée. À partir des paramètres de T2, les faits questionnés construits donneraient des questions telles que « Quels sont des résultats de la table de 3 ? » ainsi qu'un ensemble de propositions tel que $\{3, 5, 7, 9, 12\}$. Ainsi, pour offrir un algorithme de génération d'activités générique (indépendant de tout domaine didactique), il est nécessaire de posséder un objet générique dont on peut manipuler les données pour construire automatiquement des *gameplays* concrets.

Dans ce but, nous avons proposé une modélisation générique de faits questionnés (2023b). Les faits questionnés peuvent être vus comme des objets possédant une question (texte avec ou sans images), un ensemble de propositions (si la modalité de réponses est du choix parmi des propositions) associées à une valeur de véracité (correcte ou incorrecte), un ensemble de bonnes réponses (si la modalité de réponse est de la saisie) et un nombre de réponses atten-

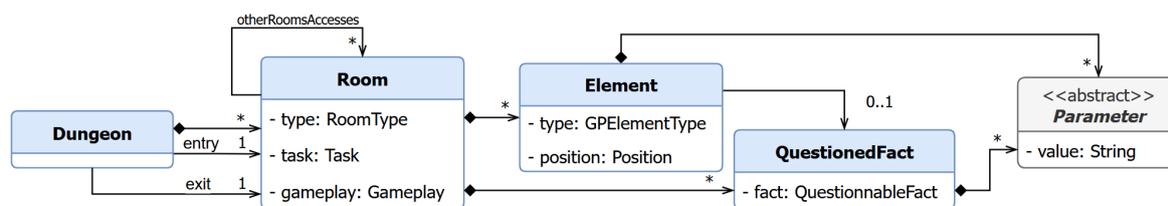


Figure 7 : Modèle conceptuel simplifié d'activités générées (Lemoine & Laforcade, 2023a, 2023b)

dues. En fonction des faits bruts et des paramètres des tâches, les paramètres des faits questionnés sont instanciés ou non. Cette modélisation permet de générer des *gameplays* concrets indépendamment du domaine didactique visé en associant les valeurs des faits questionnés aux objets de jeu correspondant aux composants décrits dans les *gameplays*. La Figure 7 présente une version simplifiée du modèle conceptuel de l'activité à générer.

4.2.4. Modèle de l'apprenant-joueur

Afin de permettre la génération d'activités adaptées à l'apprenant-joueur, il est nécessaire de connaître la progression de l'apprenant, la progression du joueur et ses préférences. Ainsi, pour connaître la progression d'un apprenant dans son parcours d'entraînement, défini par l'enseignant, ses résultats aux faits questionnés pour chaque tâche d'un couple objectif/niveau doivent être enregistrés. Cet enregistrement permet de calculer le nombre de faits qui ont été vus et travaillés ainsi que le pourcentage de succès pour chaque tâche et chaque niveau.

Dans les modèles précédents, nous avons parlé de faits bruts et de faits questionnés (questions à propos des faits bruts). Les faits questionnés possèdent, en cas de modalité de type « choix », les bonnes propositions et les mauvaises propositions. Cependant, d'après nos discussions avec les experts, il semble plus intéressant d'un point de vue didactique de constamment faire varier les réponses incorrectes. Or, du point de vue de l'apprenant, il est nécessaire de pouvoir comparer, pour une question donnée, les résultats qu'il a obtenus. Cependant, si les résultats sont enregistrés à partir des faits questionnés, la comparaison ne sera pas pertinente puisque les propositions incorrectes des faits questionnés varient à chaque génération. Deux faits questionnés pour un même fait brut seront donc considérés comme distincts. D'autre part, enregistrer les résultats à partir des faits bruts nous fait perdre l'information de la forme de question posée. En conséquence, pour comparer les résultats des apprenants sur un fait donné, il est nécessaire d'avoir une base commune qui ne varie pas (conservation du format de la question, sans les éléments qui varient).

Dans ce but, nous proposons un processus de transformation en deux étapes. Tout d'abord, les faits questionnables sont construits sur la base des faits bruts présents dans le modèle du domaine. Un fait questionnable représente une question sur un fait sans proposition incorrecte. Ces faits sont utilisés pour conserver les résultats des apprenants dans le modèle d'apprenant-joueur (e.g., les temps de réponse, les réponses données). Deuxièmement, les faits questionnés sont construits sur la base de faits questionnables (ce sont des faits questionnables avec des propositions incorrectes). La Figure 8 illustre ce processus.

Les faits questionnables sont construits à partir des tâches spécifiques aux domaines. Ainsi, le concept de fait questionnable est abstrait puisque ces faits possèdent une forme dépendante du domaine. En conséquence, notons que la génération des faits questionnables et questionnés est forcément dépendante du domaine.

De plus, pour générer un donjon correspondant à la progression de jeu (défini dans le

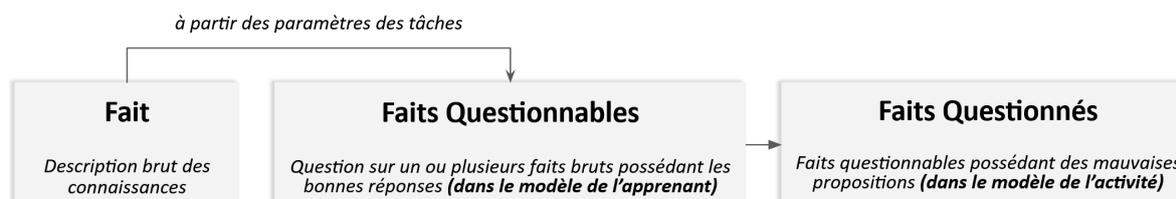


Figure 8 : Processus de transformation des faits bruts en faits questionnés

modèle de jeu), le niveau de jeu actuel du joueur doit être connu. Par exemple, si le dernier donjon réussi était de niveau 8 alors le prochain donjon généré devrait correspondre au niveau 9. Le niveau de jeu impacte le donjon en termes de nombre de salles avec et sans questions, de malédictions potentielles et de forme (labyrinthe ou linéaire).

Pour rappel, notre approche consiste à adapter les *gameplays* en fonction des préférences du joueur. En conséquence, les équipements achetés et activés ou non, doivent également être spécifiés. Ainsi, le troisième modèle à fournir au générateur est le modèle de l'apprenant-joueur qui décrit sa progression dans l'entraînement, sa progression dans le jeu et ses préférences. La Figure 9 présente une version simplifiée du modèle conceptuel de l'apprenant-joueur.

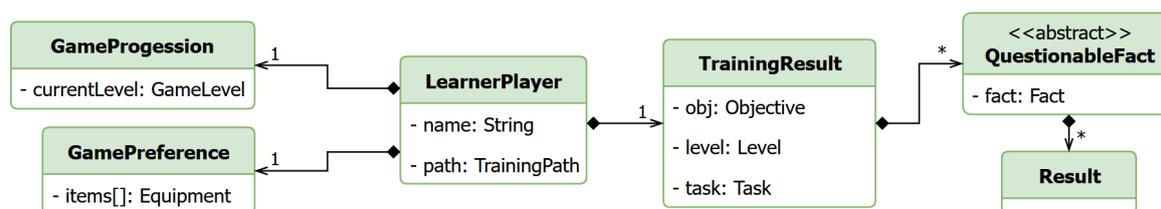


Figure 9 : Modèle conceptuel simplifié de l'apprenant-joueur (2023a)

4.2.5. Modèle de relations entre les éléments de jeu et d'entraînement

Précédemment, nous avons mentionné comment associer les éléments des faits questionnés avec des éléments de jeu correspondant à un *gameplay* (proposition de modélisation générique des faits questionnés). Cependant, certains *gameplays* ne sont pas compatibles avec certaines tâches. Par exemple, un *gameplay* consistant à orienter des statues n'a que peu de sens avec une tâche consistant à ordonner des événements chronologiquement. Ainsi, pour permettre au générateur d'associer correctement des tâches et des *gameplays*, des relations entre ces éléments doivent être définies.

L'entraînement aux connaissances déclaratives est souvent réalisé au moyen des questionnaires et des quiz. En outre, les questionnaires numériques, par rapport aux questionnaires papier, permettent des interactions avec l'utilisateur qui sont plus proches des interactions présentes dans les jeux (clic, déplacement). En conséquence, nous avons proposé une approche permettant d'associer les éléments d'entraînement et de jeu fondée sur l'utilisation des formats de questionnaires numériques comme pivot (Lemoine *et al.*, 2024a). Cette méthode systématique permet la spécification de relations interprétables par la machine entre nos types de tâche (Complétion, Identification, Identification d'Appartenance, Ordonnement) et nos catégories de *gameplay* (*SELECT*, *MOVE*, *ORIENT*, *POSITION*, *DIRECT RESPONSE*). Cette méthode permet d'obtenir des relations conditionnelles entre une catégorie de *gameplay* et une tâche. Les conditions sont un ensemble de paramètres tels que le

type de question (texte, image, texte à trous), le nombre de réponses attendues, nombre de faits questionnés et la modalité de réponse.

Dans la littérature, les relations entre les éléments éducatifs et de jeu sont rarement explicités à travers des modèles. Au contraire, elles sont souvent directement implémentées dans l’algorithme. Or, proposer un modèle interprétable par la machine permet deux choses : 1) de la modularité, c’est-à-dire la possibilité de modifier les relations sans modifier le code source, et 2) de l’extensibilité, c’est-à-dire l’ajout d’un type de tâche, le modèle peut être étendu en ajoutant les relations entre le nouveau type et les catégories de *gameplays* sans ajout de code (et inversement). En conséquence, le dernier modèle nécessaire au générateur est le modèle des relations. La Figure 10 présente le modèle conceptuel simplifié du modèle des relations entre les éléments de jeu et d’entraînement.

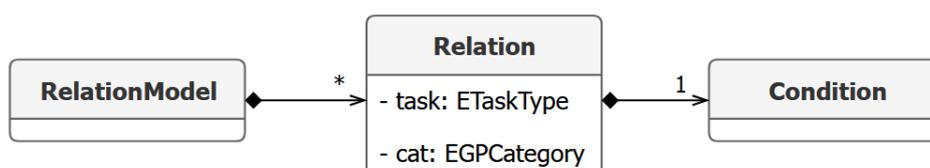


Figure 10 : Modèle conceptuel simplifié des relations (entraînement ↔ jeu) (2023a)

4.3. CADRE INFORMATIQUE : UNE APPROCHE D’INGÉNIERIE DIRIGÉE PAR LES MODÈLES

Pour permettre de guider l’implémentation de générateur d’activités, notre *framework* propose un ensemble de composants logiciels déjà existants limitant le développement de code lors de la création d’extension pour différents domaines didactique. Tout d’abord, cette section présente les méta-modèles correspondants aux modèles conceptuels précédents qui ont été traduits afin d’être instanciables et interprétables pour la génération. Ensuite, les différentes étapes de l’algorithme de génération (*générique*) d’une activité sont présentées. Enfin, l’ensemble des éléments à étendre pour construire un générateur spécifique à un domaine didactique spécifique est détaillé.

4.3.1. Méta-modèles et modèles pour la génération

Les méta-modèles et modèles sont conçus à l’aide du *framework* de modélisation d’Eclipse EMF (*Eclipse Modeling Framework*). Ces méta-modèles inter-connectés sont une représentation informatisée des modèles conceptuels présentés dans la section précédente (voir Annexe). Le méta-modèle des connaissances décrit la structure des connaissances du domaine et est à étendre en fonction du domaine didactique visé. Le méta-modèle du domaine décrit la structure des parcours d’entraînement définis par les enseignants en termes d’objectifs, niveaux, tâches, etc. Ce méta-modèle est aussi à étendre au niveau des tâches en fonction du domaine. Le méta-modèle du jeu décrit la structure de l’ensemble des types d’éléments de jeu (*gameplays*, équipements, malédictions, progression de jeu, type de salles, etc.). Le méta-modèle de l’apprenant-joueur décrit la structure de la progression d’un apprenant-joueur dans son entraînement et dans le jeu. Le méta-modèle des relations décrit la structure des conditions sous lesquelles les types de tâches génériques et les catégories de *gameplays* sont compatibles. Enfin, le méta-modèle de l’activité décrit la structure d’un donjon.

Les « instances » des méta-modèles, c’est-à-dire les modèles conformes aux méta-modèles, sont représentés sous formes de fichier XMI (*XML Metadata Interchange*) et décrivent des informations concrètes sur un domaine donné ou un jeu. La Figure 11 illustre le principe

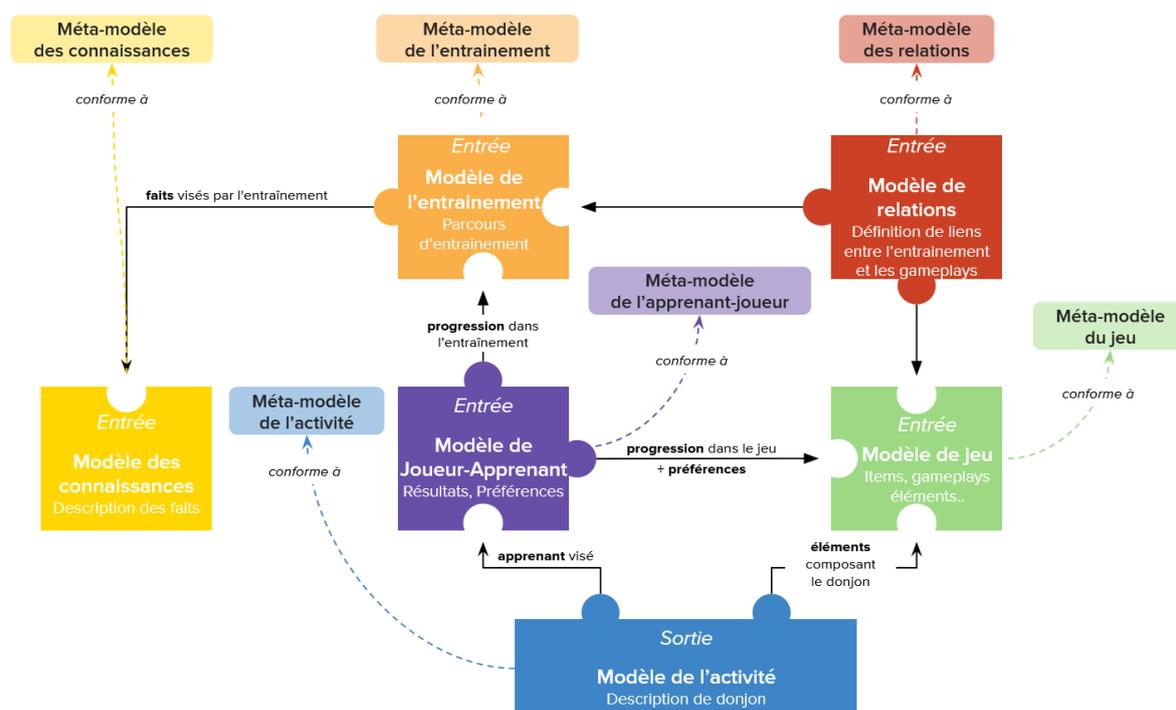


Figure 11 : Illustration des modèles interconnectés conformes aux méta-modèles impliqués dans la génération d'activités

des modèles conformes aux différents méta-modèles dans notre contexte. Les modèles des connaissances permettent de décrire les connaissances brutes à travailler (e.g., les Figures 14 et 17). Les modèles du domaine permettent de décrire des parcours d'entraînement (e.g., Figure 4). Ces instances sont dépendantes du domaine didactique. L'instance du modèle de jeu décrit l'ensemble des éléments de jeu concret disponibles pour le générateur (nous en avons défini un par défaut). Le méta-modèle de l'apprenant-joueur doit être instancié pour chaque apprenant-joueur. Les modèles d'apprenant-joueur doivent être automatiquement mis à jour après chaque niveau réalisé pour permettre à l'algorithme de génération de prendre en compte les nouveaux résultats. Cette mise à jour doit être effectuée par le jeu, en tant que composant indépendant du framework. Ensuite, nous avons « instancié » le méta-modèle des relations en fonction des résultats de la méthode d'association que nous avons précédemment proposée (Lemoine *et al.*, 2024a). Ce modèle ne doit normalement ni être ré-instancié, ni être changé, sauf en cas d'ajout de nouvelle tâche générique ou de désaccord avec notre approche. Enfin, le méta-modèle de l'activité (donjon) est « instancié » par l'algorithme de génération à partir des informations présentes dans les modèles d'entrée fournis (tous les autres).

4.3.2. Algorithme de génération d'activité

L'algorithme de génération d'une activité est implémenté en Java et s'effectue en plusieurs étapes incrémentales (similaire à Laforcade et Laghouaouta (2018) et Sehaba et Husaan (2013)). Plus précisément, l'algorithme se décompose en quatre parties découpées en neuf étapes : 1) sélection de l'ensemble des éléments du donjon (étapes une à six), 2) création de la structure (septième étape), 3) instanciation du donjon (huitième étape) et 4) transformation du donjon (neuvième étape). La Figure 12 décrit les neuf étapes de l'algorithme. Tout d'abord, le nombre de salles avec et sans question est calculé en fonction du niveau actuel du joueur dans le jeu et de la progression de jeu définie. Ensuite, un couple objec-

tif/niveau éligible du parcours d'entraînement de l'apprenant est sélectionné aléatoirement. Notons que tous les choix aléatoires réalisés par l'algorithme pourraient faire l'objet d'une heuristique donnée (e.g., stratégies données par l'enseignant pour le choix de l'objectif niveau sélectionné (El-Kechai *et al.*, 2015)). En fonction du taux d'apparition des tâches du niveau sélectionné et du nombre de salles avec question, les tâches sont choisies et ordonnées pour les salles avec question du donjon. Pour continuer, les faits questionnables du couple objectif/niveau sont générés si ce dernier n'avait jamais été sélectionné auparavant. Ensuite, le nombre nécessaire de faits questionnés pour le donjon est généré en fonction des tâches préalablement sélectionnées. Notons, que les méthodes de génération de faits appelées par l'algorithme générique sont spécifiques au domaine didactique. Par la suite, un *gameplay* de piège ou de question ainsi qu'un ensemble d'éléments de jeu à instancier compatibles avec ce *gameplay* sont sélectionnés pour chaque salle. La dernière étape de sélection consiste à choisir les malédictions du donjon (e.g., niveau dans le noir, donjon labyrinthique) en fonction du niveau de jeu atteint par l'apprenant, de la progression de jeu définie et avec une probabilité d'une sur deux.

Lorsque l'ensemble des éléments est choisi, l'algorithme génère la structure du donjon (linéaire ou labyrinthique). La génération des donjons repose sur le principe du *Grid-Based Dungeon Generator*, c'est-à-dire que 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 cellule et *grande* = 4 cellules en carré), créant un besoin de gestion des chevauchements de salles. Pour la génération de donjon linéaire, l'algorithme est fondé sur un principe de *backtrack* afin d'éviter les cul-de-sacs dus aux types de salle n'ayant pas tous les accès possibles (nord, sud, est, ouest, nord-est...) rendant la génération impossible. Ce problème ne survient pas pour les donjons labyrinthiques puisque chaque salle est éligible à chaque itération.

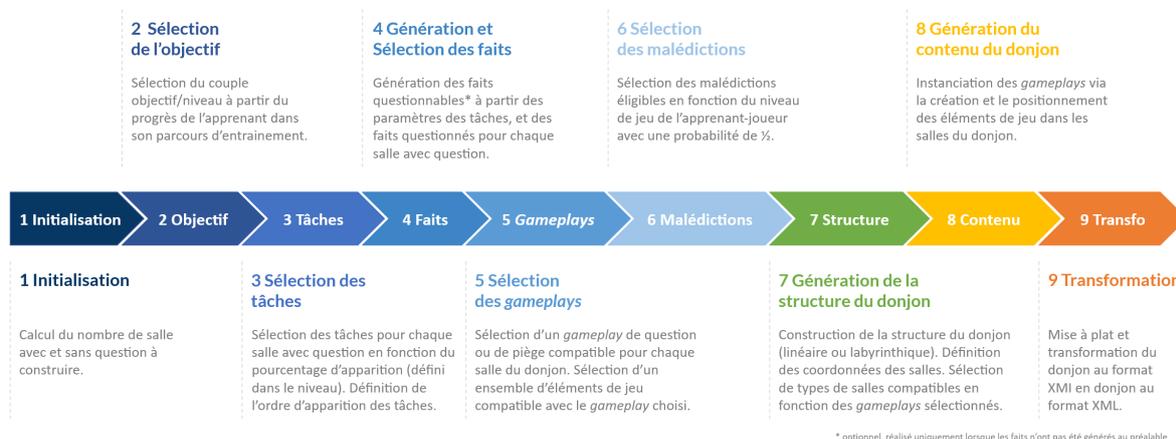


Figure 12 : Étapes de l'algorithme de génération d'activité

L'étape suivante consiste à générer le contenu du donjon en termes d'éléments de jeu. L'objectif est d'instancier correctement les éléments sélectionnés pour chaque salle du donjon en fonction des *gameplays* et des faits questionnés. Dans ce but, 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 en associant les paramètres des faits questionnés avec ceux décrits dans les *gameplays*.

Enfin, la dernière étape est une étape opérationnelle qui consiste à transformer le modèle XMI du donjon en fichier XML. Les intérêts principaux de cette transformation sont de permettre de supprimer les références vers d'autres modèles ou fichiers XMI (ce qui permet

d’aplanir le modèle) et de rendre le système plus portable. Cette transformation est réalisée à l’aide du langage *Epsilon Transformation Language* (ETL).

4.3.3. Règles d’extension

Ce *framework* doit être étendu au niveau des méta-modèles, des modèles et du code pour permettre la conception d’un générateur spécifique à un domaine. Au niveau des méta-modèles présentés, les points d’extension sont représentés par des classes abstraites (n’ayant pas déjà des sous-classes concrètes). Trois méta-modèles sont visés par cette extension : le méta-modèle des connaissances, le méta-modèle de l’entraînement et le méta-modèle de l’apprenant-joueur. Comme précédemment mentionné, les niveaux et les tâches d’entraînement peuvent posséder des paramètres spécifiques au domaine visé. Par exemple, la manière de construire les tables (opérande \times table ou table \times opérande) ne dépendent que du domaine des mathématiques. Ou bien, dans une tâche de complétion (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 événement ou date pour l’histoire). En conséquence, ces éléments doivent être modélisés en fonction du domaine. De plus, les faits bruts et questionnables ont des formes différentes en fonction du domaine didactique. 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 chaîne de caractère (événement) et une date ou période (entiers). Les faits ne peuvent donc pas être ni modélisés ni générés de manière générique.

En conséquence, les algorithmes de génération de faits questionnables et questionnés sont dépendants du domaine et doivent être implémentés au cours de l’extension du *framework*. L’algorithme de génération de faits du *framework* suit un patron de conception appelée *template-method* qui permet d’avoir une partie principale indépendante du domaine et des extensions dépendantes du domaine. La partie principale capture le squelette du code de génération, c’est-à-dire, l’ensemble des invariants relatifs à la génération. En revanche, les extensions se concentrent sur les caractéristiques orientées vers les tâches (e.g., la génération de faits questionnables ou questionnés en fonction de paramètres et de la progression actuelle d’un apprenant spécifique). Cette approche permet de soutenir la mise en œuvre de la génération de faits.

Enfin, pour pouvoir générer des activités, un modèle conforme à chaque méta-modèle doit être réalisé : modèle des connaissances décrivant les faits bruts, modèle du jeu décrivant les éléments de jeu à disposition du générateur, modèle des relations permettant au générateur de conserver la cohérence des activités, modèle du domaine décrivant le parcours d’entraînement et modèle d’apprenant-joueur. Le modèle de l’apprenant-joueur ne nécessite que les informations minimales : identifiant, parcours d’entraînement, et création des progressions (jeu et entraînement) vides, puisque ce modèle doit être mis-à-jour en fonction de résultats de l’apprenant-joueur dans le jeu. De plus, certains modèles sont définis par défaut dans le *framework* comme le modèle de jeu et le modèle de relation et sont donc réutilisables.

Pour conclure, étendre le *framework* pour concevoir un générateur nécessite de réaliser trois grandes étapes :

1. Étendre les méta-modèles en fonction du domaine visé au niveau :
 - (a) des faits bruts (*AbstractFact*);
 - (b) des faits questionnables (*AQuestionableFact*);
 - (c) des niveaux (*Level*);
 - (d) des tâches (*CompletionTask*, *IdentificationTask*, *MembershipTask*, *OrderingTask*).

2. Spécifier les modèles conformes aux méta-modèles (fichier XMI) :
 - (a) un modèle des connaissances ;
 - (b) un modèle du domaine (un parcours d’entraînement) ;
 - (c) un modèle d’apprenant (avec les informations minimales) ;
 - (d) un modèle de jeu (*optionnel*, possibilité d’utiliser celui par défaut) ;
 - (e) un modèle des relations (*optionnel*, possibilité d’utiliser celui par défaut).
3. Implémenter les générateurs de faits pour chaque tâche spécifique au domaine didactique en suivant le *template* fourni.

5. APPLICATION DU FRAMEWORK

Le *framework* proposé est une architecture produisant des générateurs. Comme défini, ce *framework* doit respecter des propriétés (voir Section 4.1). En conséquence, l’évaluation du *framework* passe dans un premier temps par la validation du respect des propriétés définies. Le *framework* doit posséder deux propriétés : FP1 permettre d’exprimer différents domaines didactiques et FP2) permettre d’exprimer la vision des enseignants sur l’entraînement des apprenants individuellement. Afin d’évaluer FP1, au moins deux extensions du *framework* doivent être conçues pour évaluer cette propriété. Nous avons actuellement développé une extension pour le domaine des tables de multiplication. Une seconde extension pour le domaine des faits de judo (prises et gestes d’arbitrage) a également été réalisée ainsi qu’une troisième extension pour le domaine des repères d’histoire-géographie du brevet des collèges (non présentée dans cet article). Afin d’évaluer FP2, au moins deux parcours d’entraînement conçus par deux enseignants différents doivent être définis pour évaluer cette propriété.

5.1. UN GÉNÉRATEUR POUR L’ENTRAÎNEMENT AUX TABLES DE MULTIPLICATION

Dans le cadre de l’étude exploratoire menée au sein du projet AdapTABLES (voir Section 3.1), nous avons défini avec des experts en mathématiques cinq tâches d’entraînement (voir Section 4.1) pour les tables de multiplication. De plus, nous avons également défini des paramètres relatifs aux niveaux de difficulté pour la construction des faits questionnables tels que la forme de construction des tables « opérande \times table ou table \times opérande », la position du symbole égal « à gauche ou à droite », l’intervalle minimal et maximal des multiplications interrogées (un enseignant peut vouloir travailler la table de trois uniquement de 1 à 5 : 3×1 , 3×2 , 3×3 , 3×4 , 3×5). À partir de ce travail, nous avons construit une extension du *framework* pour les tables de multiplication.

La Figure 13 présente l’extension du méta-modèle en question. Nous avons modélisé un type de fait brut qui représente les multiplications (*MTFact*) composé de trois entiers : la table, l’opérande et le résultat. Pour chaque tâche, nous avons créé un type de fait questionnable spécifique. En règle générale, un ou plusieurs faits questionnables sont construits à partir d’un unique fait brut, cela dépend des paramètres de constructions des tables. Cependant, dans le cas des faits pour la tâche d’identification d’appartenance (identifier les résultats d’une table), un fait questionnable est construit à partir d’un ensemble de faits bruts, en fonction des réglages de la tâche (nombre de choix attendus).

En suivant les règles d’extension, nous avons spécifié le modèle des connaissances (voir Figure 14) et avons développé les générateurs de faits pour chacune des cinq tâches. Actuellement, nous avons deux parcours d’entraînement de deux enseignantes, une de CE1 et une enseignante de CE2 pour des niveaux moyens modélisés (voir Figure 15). D’autres parcours

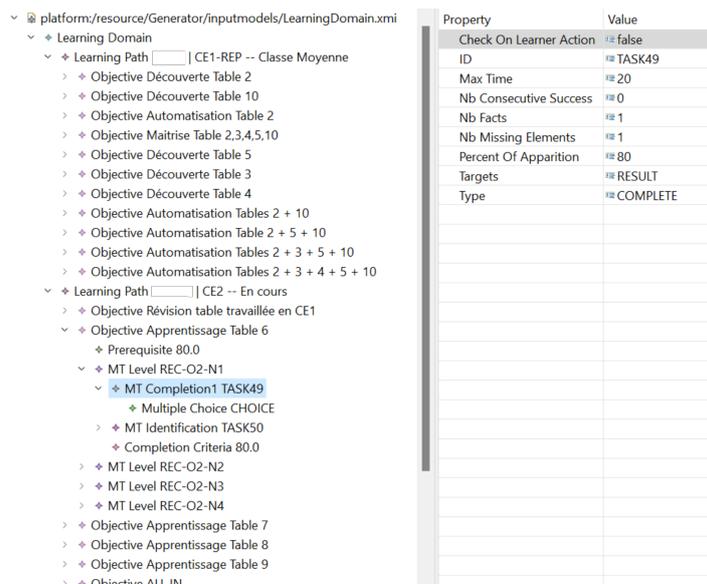


Figure 15 : Modélisation de parcours d’entrainement défini par des enseignantes en mathématiques

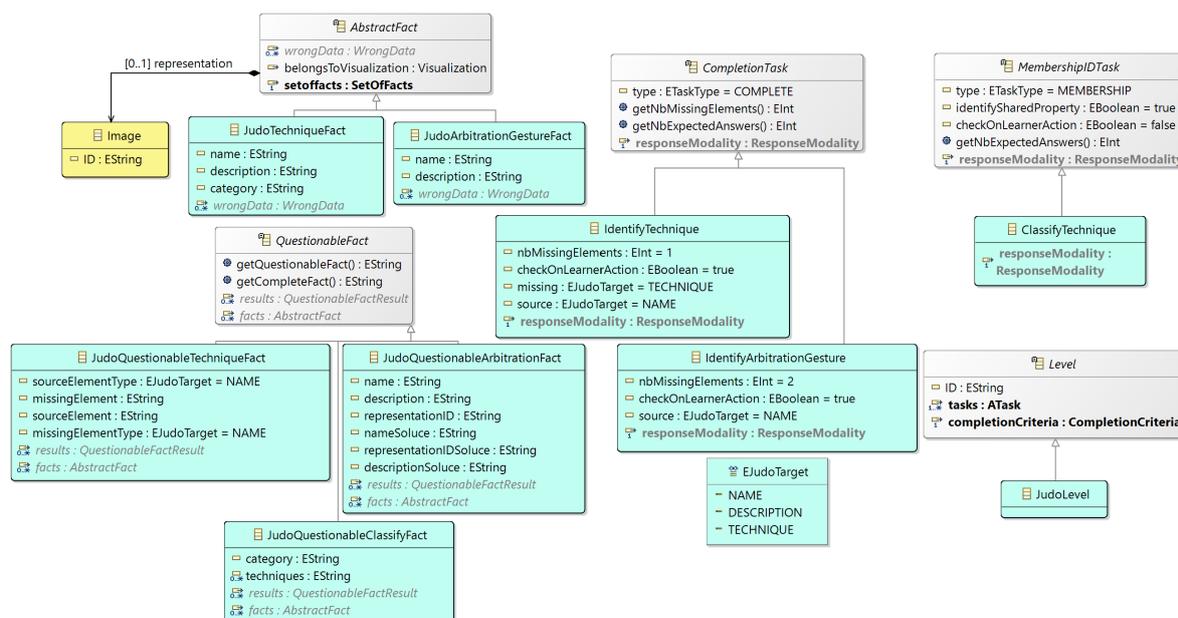


Figure 16 : Extension (couleur cyan) des méta-modèles du *framework* pour le judo

veau. Les judokas évoluent à travers un système de ceintures de couleurs (blanche, jaune, orange, verte, bleu, marron, noire) où chaque ceinture représente un niveau de compétences et de *connaissances* acquises. Les compétences à acquérir ne peuvent évidemment pas être acquises à travers notre approche, il s’agit de savoir-faire. Cependant, les connaissances à acquérir sont des connaissances déclaratives : nommer ou identifier des techniques et connaître les gestes d’arbitrages (nécessaire à tout combattant). En conséquence, nous proposons une extension du *framework* permettant de s’entraîner aux connaissances demandées au judo (cf. Figure 16).

Nous avons identifié trois tâches : 1) Identification de techniques (*IdentifyTechnique*) consistant à associer deux éléments parmi {nom de la technique, image représentative, des-

cription³}, 2) Identification des gestes d'arbitrage (*IdentifyArbitrationGesture*) consistant à reconstituer un fait composé de l'annonce de l'arbitre, la signification de son annonce et le geste associé (e.g., « HAJIME, début du combat, *main-le-long-du-corps* »), et 3) Classification de technique (*ClassifyTechnique*) consistant à identifier les techniques appartenant à une catégorie donnée. Notons qu'au judo les techniques sont classées en fonction de leur catégorie : projection de hanche, projection de jambe, immobilisation, étranglement, clé, sacrifice, etc.

Property	Value
Category	SHIME-WAZA (Techniques d'étranglement)
Description	étranglement en triangle
Name	Sankaku-Jime

Figure 17 : Modèle des connaissances pour les faits de judo

Comme pour les tables de multiplication, nous avons spécifié le modèle des connaissances (voir Figure 17) en suivant les règles d'extension et avons développé les générateurs de faits pour chacune des trois tâches. Le parcours d'entraînement étant en général le même pour l'ensemble des apprenants, nous avons défini un unique parcours d'entraînement consistant en un objectif par ceinture. Chaque objectif n'a qu'un seul niveau et a pour pré-requis le niveau de la ceinture inférieure (e.g., objectif de la ceinture jaune a pour pré-requis le niveau de l'objectif de la ceinture blanche). Enfin, chaque niveau possède trois tâches : une tâche d'identification de technique, une de geste d'arbitrage et une de classification. Évidemment, d'autres parcours peuvent être créés, ce parcours nous permet d'effectuer des tests et nous sert principalement de preuve de concept.

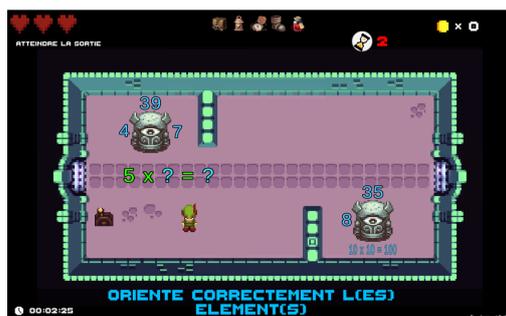
5.3. UN INTERPRÉTEUR D'ACTIVITÉS/DE NIVEAUX DE JEU

En dehors du cadre du *framework*, un interpréteur ou *player* de niveau de jeu (d'activités/donjons) a été développé. Cet interpréteur permet : 1) de retranscrire les XML générés en donjon jouables, 2) de visualiser la structure du donjon (carte) et 3) de jouer. La Figure 18 présente des captures d'écran de salles de donjon de judo et de mathématiques générées et interprétées par le *player*.

6. ÉVALUATION DU FRAMEWORK

Comme défini, les générateurs produits par le *framework* doivent respecter des propriétés (voir Section 4.1) :

3. Une description est souvent donnée par les enseignants pour décrire les techniques « Ippon-seoi-nage : projection d'épaule par un côté »



(a) Complétion un fait



(b) Reconstitution de fait



(c) Identification de résultats



(d) Identification de technique



(e) Identification de geste d'arbitrage



(f) Classification de technique

Figure 18 : Captures d'écran de salles de donjon interprétées par le *player*

- GP1) les activités générées doivent être adaptées à l'apprenant ;
- GP2) les activités générées doivent être adaptées aux préférences du joueur ;
- GP3) les activités générées doivent être variées en termes d'éléments éducatifs et d'éléments de jeu.

L'évaluation de ces propriétés peut être réalisée par des tests systèmes automatisés (JUnit) consistant à vérifier le respect d'un ensemble d'intentions données.

6.1. VALIDATION DES ACTIVITÉS GÉNÉRÉES : TESTS SYSTÈMES

Les générateurs construits sont spécifiques à un domaine didactique. En conséquence, vérifier que les activités produites sont adaptées et variées, consiste à vérifier à l'aide de tests systèmes que les activités respectent les propriétés attendues (voir Section 4.1). Les tests systèmes utilisent un générateur spécifique à un domaine didactique donné. Dans notre contexte, les tests systèmes ont été réalisés à partir du générateur des mathématiques. Cependant, l'algorithme étant générique (seule la génération de fait est spécifique au domaine), les résultats peuvent être généralisés (e.g., l'algorithme de sélection d'objectif ne change pas

d'un domaine à un autre).

6.1.1. Adaptation des activités 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'entraînement associé à l'apprenant. La conformité de l'activité repose sur deux niveaux d'analyse : A) sélection du couple objectif/niveau à travailler et B) tâches présentes dans le donjon en fonction du niveau sélectionné, de la progression de l'apprenant et du pourcentage d'apparition souhaité.

A) SÉLECTION COUPLE OBJECTIF/NIVEAU. L'algorithme de génération doit choisir un couple objectif/niveau éligible (plusieurs couples objectif/niveau peuvent être éligibles en même temps) en fonction de la description du parcours d'entraînement et de la progression de l'apprenant dans ce parcours. Pour évaluer l'algorithme de sélection d'un couple objectif/niveau, nous avons défini un parcours d'entraînement et un modèle d'apprenant fictif incluant les différents cas limites à tester. À partir de ces modèles, nous avons créé une méthode de tests pour chacune des intentions suivantes :

- le couple objectif/niveau sélectionné est éligible ;
- pour 150 générations, tous les couples objectif/niveau éligibles sont apparus au moins une fois ;
- pour 150 générations, aucun des couples objectif/niveau non éligibles n'est apparu ;
- les niveaux dont le pourcentage de faits rencontrés et le pourcentage de succès sont atteints sont considérés comme inéligibles (et donc n'apparaissent jamais dans les donjons).

B) RÉPARTITION DES TÂCHES. L'algorithme de génération doit répartir les tâches en fonction : 1) de leur taux d'apparition, 2) du nombre de salles du donjon dépendant de la progression du joueur dans le jeu (niveau de jeu actuel du joueur). Par exemple, une tâche ayant un taux de 20% devra apparaître dans 2 salles d'un donjon de taille 10. Notons que lorsque la taille d'un donjon et le taux d'apparition d'une tâche sont petits (e.g., cinq salles et 5% d'apparition), cette tâche peut ne pas apparaître dans le donjon. En conséquence, des tests doivent être effectués sur deux niveaux de jeu différents (tailles de donjon différentes). Lorsqu'une tâche est terminée, l'algorithme répartit son pourcentage d'apparition proportionnellement aux pourcentages d'apparition des tâches restantes : soit T1[50%], T2[20%], T3[20%], T4[10%] où T1 est achevée, alors les pourcentages des autres tâches deviennent T2[40%], T3[40%], T4[20%].

Pour évaluer la répartition des tâches, nous avons défini deux parcours d'entraînement de test (voir Figure 19). Le premier est composé de 5 tâches avec un taux d'apparition identiques (20%), c'est-à-dire qu'aucune tâche ne prédomine sur les autres. Le second est composé de 6 tâches avec des taux d'apparition différents, c'est-à-dire que certaines tâches sont prédominantes. Quatre situations critiques (cas limites) sont à vérifier : 1) l'apprenant n'a pas commencé l'entraînement, toutes les tâches ont un taux de succès et de faits rencontrés à 0% (toutes les tâches apparaissent proportionnellement à leur pourcentage d'apparition dans le donjon) ; 2) toutes les tâches ont été démarrées, mais aucune n'a été terminée (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 (les deux pourcentages de cette tâche ont atteint 100% et la tâche n'apparaît plus dans le donjon) ; 4) toutes les tâches, sauf une, ont été terminées (seule la tâche non terminée apparaît dans le donjon). Pour chacun des cas limite et chacun des parcours d'entraînement, nous avons spécifié deux modèles d'apprenant-joueur : un modèle dans lequel le joueur est au niveau de jeu 1

(cinq salles avec questions) et un modèle dans lequel le joueur est au niveau de jeu 16 (20 salles avec questions). Ensuite, pour chaque modèle d'apprenant-joueur, une méthode de test vérifiant que le donjon généré possède le bon nombre de salles par tâche a été implémentée.

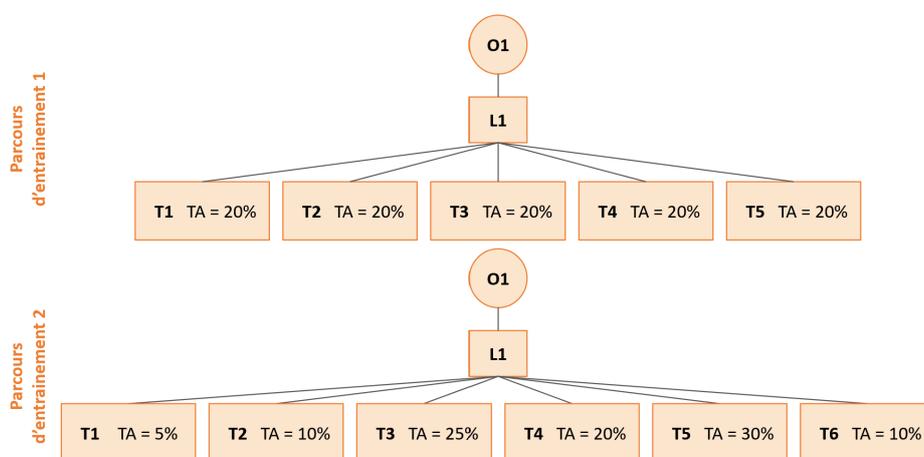


Figure 19 : Parcours d'entraînement fictifs permettant de tester l'algorithme de répartition des tâches

6.1.2. Adaptation des activités au joueur

Tout comme l'évaluation de l'adaptation à l'apprenant, l'adaptation au joueur consiste à vérifier que les activités générées sont conformes aux prédictions déductibles d'une analyse des préférences de jeu associées au joueur. Les préférences de jeu du joueur, dans notre contexte, sont décrites au travers d'équipement achetables et activables. Ces équipements permettent de débloquer des capacités et donc de déverrouiller de nouveaux *gameplays*. Un joueur peut activer ou désactiver l'équipement, bloquant alors l'apparition de cette capacité et des *gameplays* associés en fonction de ses préférences de jeu. Bien sûr, des *gameplays* par défaut, ou *gameplays* de base, sont définis et non désactivables, sinon le jeu ne serait pas jouable. En conséquence, nous avons implémenté des méthodes de tests représentant les quatre cas limites :

- l'apprenant-joueur n'a effectué aucun achat, alors seuls les *gameplays* par défaut sont présents dans les donjons ;
- l'apprenant-joueur a acheté tous les équipements possibles, mais n'en a activé aucun, alors seuls les *gameplays* par défauts sont présents dans les donjons ;
- l'apprenant-joueur a acheté et activé tous les équipements possibles, alors tous les *gameplays* apparaissent au moins une fois dans un des donjons générés ;
- l'apprenant-joueur a acheté et activé certains équipements, les *gameplays* associés aux équipements achetés et activés apparaissent au moins une fois dans un donjon à un moment donné et les autres n'apparaissent jamais.

6.1.3. Variété des activités

La variété des donjons repose à la fois sur le contenu d'entraînement et sur le contenu de jeu. Nous nous intéressons à la variété des activités au niveau : A) de la structure/répartition des donjons et B) des éléments dans les donjons.

A) STRUCTURE/RÉPARTITION DES DONJONS. Dans les *Roguelites*, deux donjons n'ont jamais la même forme (salles, objets). Tout est organisé différemment à chaque fois. L'interpréteur de donjon (voir Section 5.3) permet de visualiser les donjons sans avoir à les explorer. Cet interpréteur permet de visualiser une carte dans laquelle chaque salle est colorée en fonction de la tâche associée et où les accès entre chaque salle sont définis par un petit trait (permet de différencier les donjons labyrinthiques des donjons linéaires). Par conséquent, une première évaluation s'appuie sur une évaluation manuelle consistant à générer, pour un même niveau, plusieurs donjons et à comparer graphiquement leur carte. La Figure 20 présente quatre cartes de donjons générés pour un même objectif ou niveau donné et un même apprenant joueur (avec une progression n'ayant pas changé). On peut facilement remarquer que ces donjons présentent différentes structures et sont répartis de façon non identique.

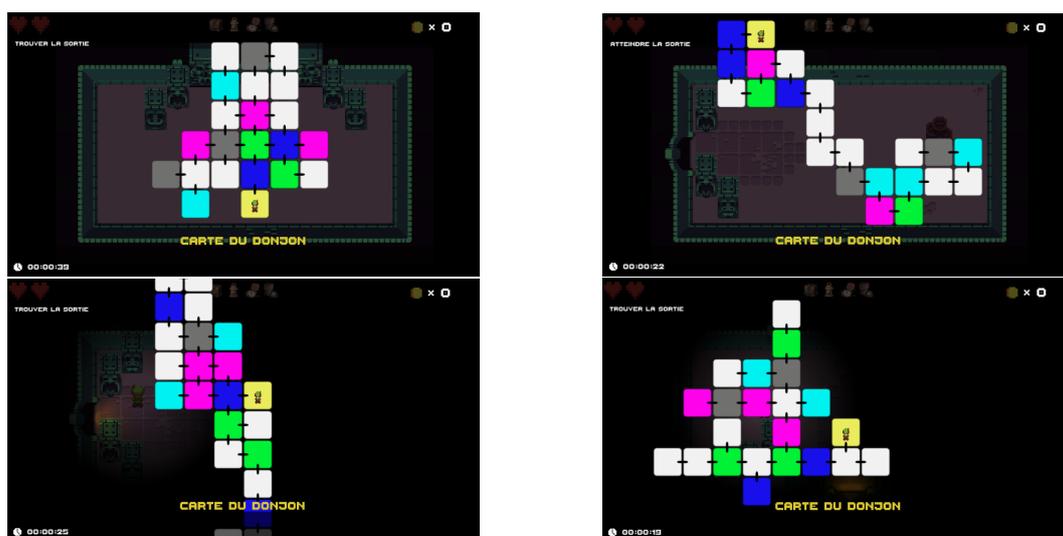


Figure 20 : Cartes de quatre donjons générés pour un même couple objectif/niveau et un même apprenant-joueur

B) ÉLÉMENTS DES DONJONS. La variété des éléments des donjons peut être considérée à différents niveaux : 1) variété des *gameplays* sélectionnés pour chaque tâche; 2) variété des éléments pour chaque *gameplay*; 3) variété des types de salle choisis (*RoomType*); 4) variété de la position des éléments dans les salles; 5) variété des mauvais choix des faits questionnés. La variété des types des salles et de la position des éléments est dépendante de la modélisation réalisée dans le modèle de jeu. De plus, la variété des mauvais choix des faits questionnés dépend du domaine didactique visé puisque la méthode créée pour le générer fait partie de l'extension. La variété des *gameplays* sélectionnés pour chaque tâche dépend de deux choses : 1) la variété de *gameplays* décrit dans le modèle de jeu et 2) les achats effectués par le joueur. Pour chacune des capacités (capacités par défaut et capacités verrouillables), différents *gameplays* et variantes de *gameplays* ont été définis (e.g., une variante de "pousser des éléments vers la droite" est de "pousser des éléments vers la gauche") afin que chaque tâche apparaisse avec au moins deux *gameplays* différents ou deux variantes de *gameplays*. Cet aspect (la sélection d'au moins deux modes de jeu différents pour chaque tâche sur 100 générations de donjons) a été évalué pour un joueur n'ayant rien acheté et pour un joueur ayant tout acheté et tout activé. Enfin, la variété des éléments pour chaque *gameplay* dépend majoritairement de la variété des éléments de jeu décrit dans le modèle de jeu. Nous n'évaluons pas formellement ce critère pour le moment, cependant, l'approche semble

prometteuse puisque nous trouvons des variantes d'un même gameplay en termes d'éléments de jeu dans les donjons. Par exemple, nous avons une capacité nommée *CATCHABLE* pour laquelle nous avons deux éléments de jeu associés : le lapin et la vache. Le *gameplay* associé à la capacité *CATCHABLE* apparaît dans les donjons avec les deux éléments de jeu différents (voir Figure 21).

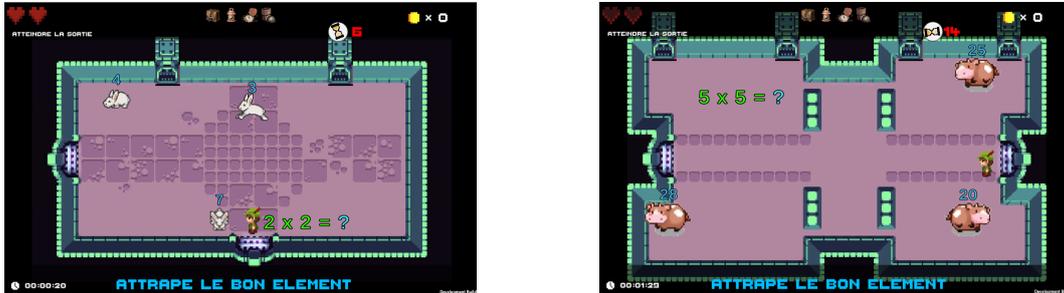


Figure 21 : Deux variantes d'un même *gameplay*

6.2. VÉRIFICATION DES MODÈLES : VALIDATION DE CONTRAINTES STATIQUES

Une autre étape de vérification du *framework* consiste à valider les modèles qu'il utilise. Si de nombreuses informations sont capturées par les méta-modèles, ces derniers ne permettent pas de capturer l'ensemble des propriétés sémantiques voulues pour un domaine donné. Par exemple, dans notre modèle du domaine actuel (voir Figure 24 en annexe) rien ne nous empêche de définir qu'un objectif a pour pré-requis son propre niveau. Ce qui sémantiquement n'a aucun sens. La validation de contraintes statiques sur les modèles est une méthode semi-formelle de vérification automatique de modèles consistant à définir un ensemble de contraintes et à valider ou non les modèles en fonction de ces contraintes. Pour garantir la bonne sémantique des modèles (génériques) fournis en entrée des générateurs, nous avons défini un ensemble de contraintes sémantiques sur les modèles à l'aide du langage *Epsilon Validation Language* (EVL).

```

/*
Checks if every SetOfFacts of an Objective, belongs to the
corresponding Knowledge of the objective's path.
*/
constraint objectiveFactsBelongsToPathKnowledge {
  check {
    return self.objectives.forAll(obj |
      obj.setoffacts.forAll(facts |
        self.knowledge.knowledgefacts.contains(facts)));
  }
  message {
    return "SetOfFacts does not belong to the correct knowledge";
  }
}

```

Figure 22 : Exemple de règle de validation de modèles écrite en EVL vérifiant si les faits associés à un objectif appartiennent au modèle de connaissances associé au parcours concerné

Par exemple, les règles définies pour la validation du modèle du domaine sont :

- les faits d'un objectif appartiennent aux connaissances associées au parcours d'entraînement (la Figure 22 illustre cette règle en EVL);
- le niveau (*requiredLevel*) d'un pré-requis d'un objectif O, n'est pas un niveau de l'objectif O;

- les pourcentages des succès et de faits rencontrés (*successPercent* et *encountersPercent*) des pré-requis et des critères de complétion ont des valeurs entre 0 et 100);
- les pourcentages de succès et de faits rencontrés des pré-requis doivent être inférieurs ou égaux aux pourcentages du critère de complétion (sinon ce n'est pas atteignable).

7. CONCLUSION & PERSPECTIVES

Cet article présente un *framework* guidant la conception et l'implémentation de générateurs d'activités d'entraînement de jeu de type Roguelite aux connaissances déclaratives. Le *framework* est une architecture logicielle extensible fondée sur une approche d'Ingénierie Dirigée par les Modèles, c'est-à-dire utilisant des modèles et méta-modèles interprétables par la machine. Ce *framework* permet de faciliter l'implémentation de générateurs d'activités, par des ingénieurs ou développeurs, en limitant le développement nécessaire aux informations reliées aux connaissances déclaratives du domaine didactique. Ainsi, ce *framework* propose un ensemble d'algorithmes, modèles, et méta-modèles déjà existants à étendre qui seront donc réutilisés pour spécifier un générateur d'activités.

Pour évaluer, la généralité et l'extensibilité du *framework*, trois générateurs ont été produits à partir de celui-ci : un générateur pour l'entraînement aux tables de multiplication, un second pour les techniques et les gestes d'arbitrages de judo, et un troisième pour les repères d'histoire-géographie au programme du brevet des collèges. Les propriétés d'adaptation à l'apprenant-joueur et de variété des activités ont été évaluées à travers des tests automatisés.

À court terme, nous souhaitons automatiser les tests au niveau des variantes d'éléments pour un *gameplay* donné. Pour rappel, des tests permettant de vérifier la variété de *gameplays* proposés pour une tâche d'entraînement donnée ont été implémentés. Ces tests complémentaires consisteraient à vérifier que, pour un même *gameplay* sélectionné, différents éléments de jeu ayant les capacités référencées dans ce *gameplay* sont choisis par l'algorithme de génération.

Également, à court-terme, nous préparons l'expérimentation directe du *framework* par des experts et ingénieurs maîtrisant l'IDM. Ils auront à étendre le *framework* à un domaine que nous leur fournissons : spécifications non informatiques des faits à entraîner, les tâches d'entraînement souhaitées, etc. Il s'agira d'évaluer la *documentation* guidant cette *extension* ainsi que la facilité/difficulté de l'extensibilité du *framework* (*effort* de modélisation et de codage). L'IDM étant un cadre théorique et pratique non maîtrisé par l'ensemble des ingénieurs, une perspective possible, à plus long terme, serait d'instrumenter les activités de l'expert en proposant un environnement logiciel interactif pour assister la création d'extensions pour différents domaines didactiques.

À moyen terme, nous aimerions également prendre en compte d'autres composantes pédagogiques de l'entraînement qui pourraient faire l'objet d'adaptations à l'apprenant. En effet, actuellement, soit la génération ne précise pas certaines informations qui sont laissées à l'appréciation du jeu (e.g., le message lorsque la réponse donnée est correcte ou mauvaise), soit la génération réalise des choix au hasard (e.g., le choix du couple objectif/niveau à entraîner parmi ceux éligibles). Ces choix arbitraires, ou laissés au hasard, peuvent faire l'objet de nouvelles adaptations. Par exemple, les *feedbacks* pourraient être personnalisés en fonction de préférences de l'élève ou bien de l'enseignant, ils pourraient également mieux prendre en compte le contexte de réponse (temps de réponse, erreurs passées pour le même fait, etc.). De plus, le choix pour chaque niveau de jeu du couple objectif et niveau d'entraînement visés pourrait s'appuyer sur des stratégies pédagogiques précisées par l'enseignant pour chaque élève (e.g., maîtriser d'abord un objectif avant de passer à un autre).

À plus long terme, nous souhaiterions également nous intéresser à la génération d'activités de jeu pour d'autres types de connaissances : les connaissances procédurales. Ces connaissances plus complexes nécessitent d'entraîner la réalisation d'actions spécifiques dans un contexte didactique donné. Une génération adaptée et variée nécessiterait de s'intéresser au besoin de l'entraînement de ces connaissances, mais également à d'autres genres de jeu pouvant répondre à ces besoins.

RÉFÉRENCES

- Amory, A. (2007). Game object model version II : a theoretical framework for educational game development. *Education Tech Research Dev*, 55(1), 51-77. <https://doi.org/10.1007/s11423-006-9001-x>
- Bakkes, S., Tan, C. T., et Pisan, Y. (2012). Personalised gaming : a motivation and overview of literature. *Proceedings of The 8th Australasian Conference on Interactive Entertainment : Playing the System*, 1-10. <https://doi.org/10.1145/2336727.2336731>
- Bezza, A., Balla, A., et Marir, F. (2013). An approach for personalizing learning content in e-learning systems : A review. *2013 Second International Conference on E-Learning and E-Technologies in Education (ICEEE)*, 218-223. <https://doi.org/10.1109/ICeLeTE.2013.6644377>
- Bontchev, B. P., Terzieva, V., et Paunova-Hubenova, E. (2021). Personalization of serious games for learning. *ITSE*, 18(1), 50-68. <https://doi.org/10.1108/ITSE-05-2020-0069>
- Brame, C. J., et Biel, R. (2015). Test-enhanced learning : the potential for testing to promote greater learning in undergraduate science courses. *LSE*, 14(2). <https://doi.org/10.1187/cbe.14-11-0208>
- Callies, S., Sola, N., Beaudry, E., et Basque, J. (2015). An empirical evaluation of a serious simulation game architecture for automatic adaptation. R. Munkvold & L. Kolas, *Proceedings of the 9th European Conference on Games Based Learning (ECGBL 2015)*, 107-116.
- Carpentier, K., et Lourdeaux, D. (2014). Generation of learning situations according to the learner's profile within a virtual environment. Dans J. Filipe et A. Fred (dir.), *Agents and Artificial Intelligence* (p. 245-260, T. 449). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-44440-5_15
- Carvalho, M. B., Bellotti, F., Berta, R., De Gloria, A., Sedano, C. I., Hauge, J. B., Hu, J., et Rauterberg, M. (2015). An activity theory-based model for serious games analysis and conceptual design. *Computers & Education*, 87, 166-181. <https://doi.org/10.1016/j.compedu.2015.03.023>
- Codish, D., et Ravid, G. (2015). Adaptive approach for gamification optimization. *2014 IEEE/ACM 7th International Conference on Utility and Cloud Computing, UCC 2014*, 609-610. <https://doi.org/10.1109/UCC.2014.94>
- Dias, T. (2018). *Enseigner les mathématiques à l'école*. Magnard.
- Diwan, C., Srinivasa, S., et Ram, P. (2019). Automatic generation of coherent learning pathways for open educational resources. Dans M. Scheffel, J. Broisin, V. Pammer-Schindler, A. Ioannou et J. Schneider (dir.), *Transforming Learning with Meaningful Technologies* (p. 321-334, T. 11722). Springer International Publishing. https://doi.org/10.1007/978-3-030-29736-7_24
- Djaouti, D., Alvarez, J., Jessel, J.-P., Methel, G., et Molinier, P. (2008). A gameplay definition through videogame classification. *International Journal of Computer Games Technology*, 1-7. <https://doi.org/10.1155/2008/470350>

- Dormans, J., et Bakkes, S. (2011). Generating missions and spaces for adaptable play experiences. *IEEE Trans. Comput. Intell. AI Games*, 3(3), 216-228. <https://doi.org/10.1109/TCIAIG.2011.2149523>
- El-Kechai, N., Melero, J., et Labat, J.-M. (2015). Adaptation de serious games selon la stratégie choisie par l'enseignant : approche fondée sur la Competence-based Knowledge Space Theory. *7ème Conférence sur les Environnements Informatiques pour l'Apprentissage Humain (EIAH 2015)*, 294-305.
- Grant, P., et Basye, D. (2014). *Personalized learning : a guide for engaging students with technology*. International Society for Technology in Education.
- Holohan, E., Melia, M., McMullen, D., et Pahl, C. (2006). The generation of e-learning exercise problems from subject ontologies. *Sixth IEEE International Conference on Advanced Learning Technologies (ICALT'06)*, 967-969. <https://doi.org/10.1109/ICALT.2006.1652605>
- Hunicke, R., LeBlanc, M., et Zubek, R. (2004). MDA : a formal approach to game design and game research. *Proceedings of the AAAI Workshop on Challenges in Game AI*, 4.
- Ismail, H., et Belkhouche, B. (2018). A reusable software architecture for personalized learning systems. *2018 International Conference on Innovations in Information Technology (IIT)*, 105-110. <https://doi.org/10.1109/INNOVATIONS.2018.8605997>
- Jézéquel, J.-M., Combemale, B., et Vojtisek, D. (2012, février). *Ingénierie dirigée par les modèles : des concepts à la pratique...* Ellipses.
- Junior, R., et Silva, F. (2021). Redefining the MDA framework—the pursuit of a game design ontology. *Information*, 12(10). <https://doi.org/10.3390/info12100395>
- Kent, S. (2002). Model driven engineering. *Integrated Formal Methods*, 286-298.
- Kim, J. W., Ritter, F. E., et Koubek, R. J. (2013). An integrated theory for improved skill acquisition and retention in the three stages of learning. *Theoretical Issues in Ergonomics Science*, 14(1), 22-37. <https://doi.org/10.1080/1464536X.2011.573008>
- Laforcade, P., et Laghouaouta, Y. (2018). Generation of adapted learning game scenarios : a model-driven engineering approach. Dans B. M. McLaren, R. Reilly, S. Zvacek et J. Uhomobhi (dir.), *Computer Supported Education - 10th International Conference, CSEDU 2018, Funchal, Madeira, Portugal, March 15-17, 2018, Revised Selected Papers* (p. 95-116, T. 1022). Springer. https://doi.org/10.1007/978-3-030-21151-6_6
- Laforcade, P., Mottier, E., Jolivet, S., et Lemoine, B. (2022). Expressing adaptations to take into account in generator-based exercisers : an exploratory study about multiplication facts. *14th International Conference on Computer Supported Education*. <https://doi.org/10.5220/0011033100003182>
- Lemoine, B., et Laforcade, P. (2023a). Generator of personalised training games activities : a conceptual design approach. Dans P. Dondio, M. Rocha, A. Brennan, A. Schönbohm, F. de Rosa, A. Koskinen et F. Bellotti (dir.), *Games and Learning Alliance - 12th International Conference, GALA 2023, Dublin, Ireland, November 29 - December 1, 2023, Proceedings* (p. 321-331, T. 14475). Springer. https://doi.org/10.1007/978-3-031-49065-1_31
- Lemoine, B., et Laforcade, P. (2023b). Mapping facts to concrete game elements for generation purposes : a conceptual approach. Dans P. Dondio, M. Rocha, A. Brennan, A. Schönbohm, F. de Rosa, A. Koskinen et F. Bellotti (dir.), *Games and Learning Alliance - 12th International Conference, GALA 2023, Dublin, Ireland, November 29 - December 1, 2023, Proceedings* (p. 342-352, T. 14475). Springer. https://doi.org/10.1007/978-3-031-49065-1_33

- Lemoine, B., Laforcade, P., et George, S. (2023). Un Framework de conception pour des générateurs d'activités de jeu variées et adaptées. *11ème Conférence Sur Les Environnements Informatiques Pour l'Apprentissage Humain*, 88-99.
- Lemoine, B., Laforcade, P., et George, S. (2024a). An approach for mapping declarative knowledge training task types to gameplay categories. Dans B. M. McLaren, J. Uhomobhi, J. Jovanovic et I.-A. Chounta (dir.), *Computer Supported Education* (p. 47-68). Springer Nature Switzerland.
- Lemoine, B., Laforcade, P., et George, S. (2024b). Designing declarative knowledge training games : an analysis framework based on the roguelite genre. Dans B. M. McLaren, J. Uhomobhi, J. Jovanovic et I.-A. Chounta (dir.), *Computer Supported Education* (p. 69-92). Springer Nature Switzerland.
- Marne, B., Carron, T., Labat, J.-M., et Marfisi-Schottman, I. (2013). MoPPLiq : a model for pedagogical adaptation of serious game scenarios. *IEEE 13th International Conference on Advanced Learning Technologies, ICALT 2013, Beijing, China, July 15-18, 2013*, 291-293. <https://doi.org/10.1109/ICALT.2013.90>
- Marty, J.-C., et Carron, T. (2011). Hints for improving motivation in game-based learning environments. Dans *Handbook of Research on Improving Learning and Motivation through Educational Games : Multidisciplinary Approaches* (p. 530-549). IGI Global.
- Monterrat, B., Yessad, A., Bouchet, F., Lavoué, É., et Luengo, V. (2017). MAGAM : a multi-aspect generic adaptation model for learning environments. Dans É. Lavoué, H. Drachler, K. Verbert, J. Broisin et M. Pérez-Sanagustín (dir.), *Data Driven Approaches in Digital Education* (p. 139-152, T. 10474). Springer International Publishing. https://doi.org/10.1007/978-3-319-66610-5_11
- Natkin, S., Yan, C., Jumpertz, S., et Marquet, B. (2007). Creating multiplayer ubiquitous games using an adaptive narration model based on a user's model. Dans A. Baba (dir.), *Proceedings of the 2007 DiGRA International Conference : Situated Play, DiGRA 2007, Tokyo, Japan, September 24-28, 2007*. Digital Games Research Association.
- Plass, J. L., et Pawar, S. (2020). Toward a taxonomy of adaptivity for learning. *Journal of Research on Technology in Education*, 52(3), 275-300. <https://doi.org/10.1080/15391523.2020.1719943>
- Prensky, M. (2005). *Computer games and learning : digital game-based learning*. Handbook of Computer Game Studies.
- Roediger, H. L., et Pyc, M. A. (2012). Inexpensive techniques to improve education : applying cognitive psychology to enhance educational practice. *Journal of Applied Research in Memory and Cognition*, 1(4), 242-248. <https://doi.org/10.1016/j.jarmac.2012.09.002>
- Roepke, R., Drury, V., Schroeder, U., et Meyer, U. (2021). A modular architecture for personalized learning content in anti-phishing learning games. *Software Engineering (Satellite Events)*.
- Sehaba, K., et Hussaan, A. M. (2013). GOALS : generator of adaptive learning scenarios. *IJLT*, 8(3), 224. <https://doi.org/10.1504/IJLT.2013.057061>
- Sina, S., Rosenfeld, A., et Kraus, S. (2014). Generating content for scenario-based serious-games using crowdsourcing. *Proceedings of the National Conference on Artificial Intelligence*, 1, 522-529.
- Smith, R. P. (1981). Boredom : a review. *Human factors*, 23(3), 329-340.
- Streicher, A., et Smeddinck, J. D. (2016). Personalized and adaptive serious games. Dans R. Dörner, S. Göbel, M. Kickmeier-Rust, M. Masuch et K. Zweig (dir.), *Entertain-*

ment Computing and Serious Games (p. 332-377, T. 9970). Springer International Publishing. https://doi.org/10.1007/978-3-319-46152-6_14

Tchounikine, P., Mørch, A. I., et Bannon, L. J. (2009). A computer science perspective on technology-enhanced learning research. Dans *Technology-Enhanced Learning* (p. 275-288). Springer Netherlands. https://doi.org/10.1007/978-1-4020-9827-7_16

Vandewaetere, M., Desmet, P., et Clarebout, G. (2011). The contribution of learner characteristics in the development of computer-based adaptive learning environments. *Computers in Human Behavior*, 27(1), 118-130. <https://doi.org/10.1016/j.chb.2010.07.038>

ANNEXE : MÉTA-MODÈLES POUR LA GÉNÉRATION

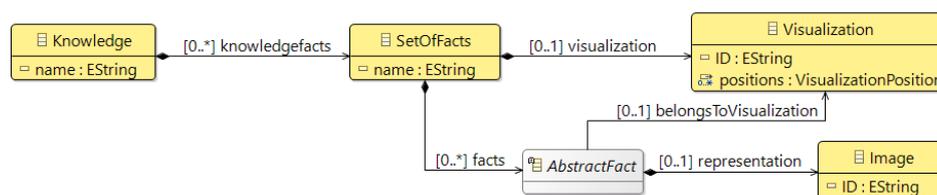


Figure 23 : Méta-modèle des connaissances

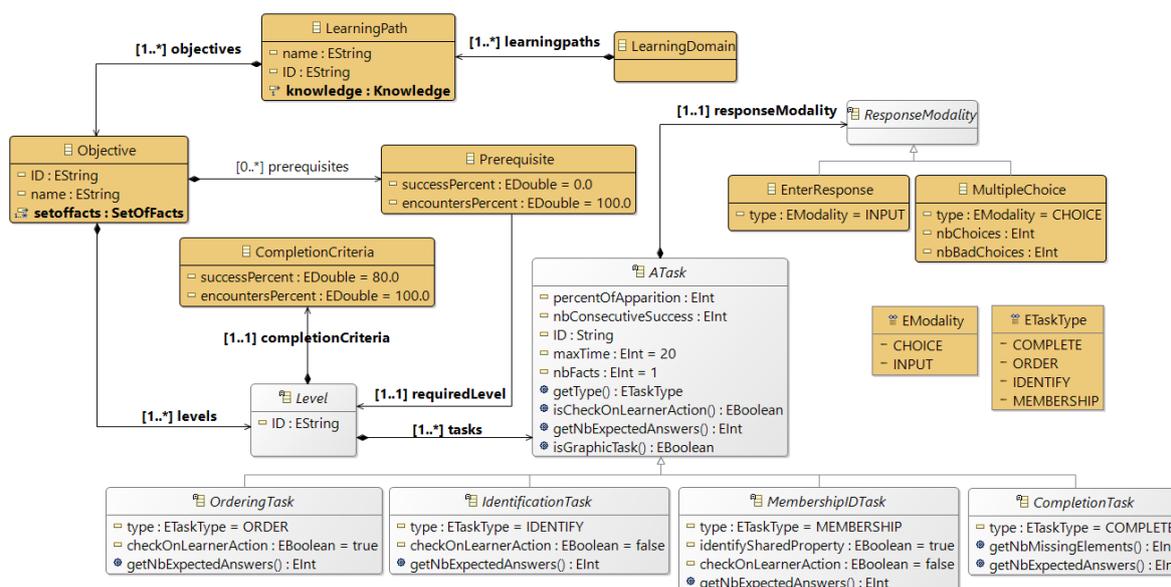


Figure 24 : Méta-modèle de l'entraînement

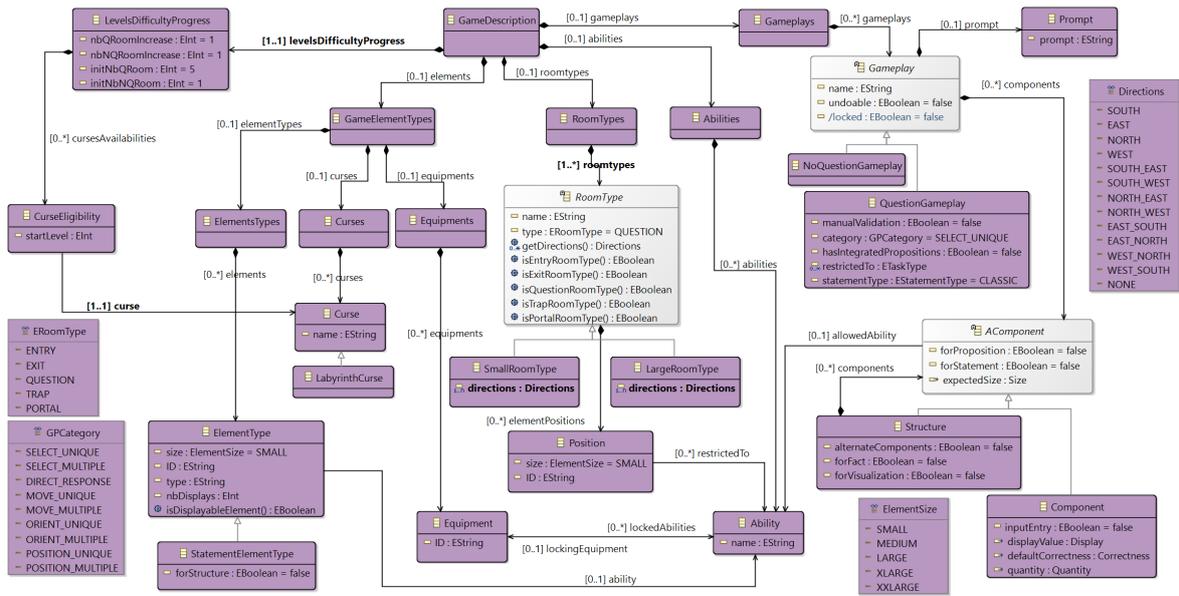


Figure 25 : Méta-modèle du jeu

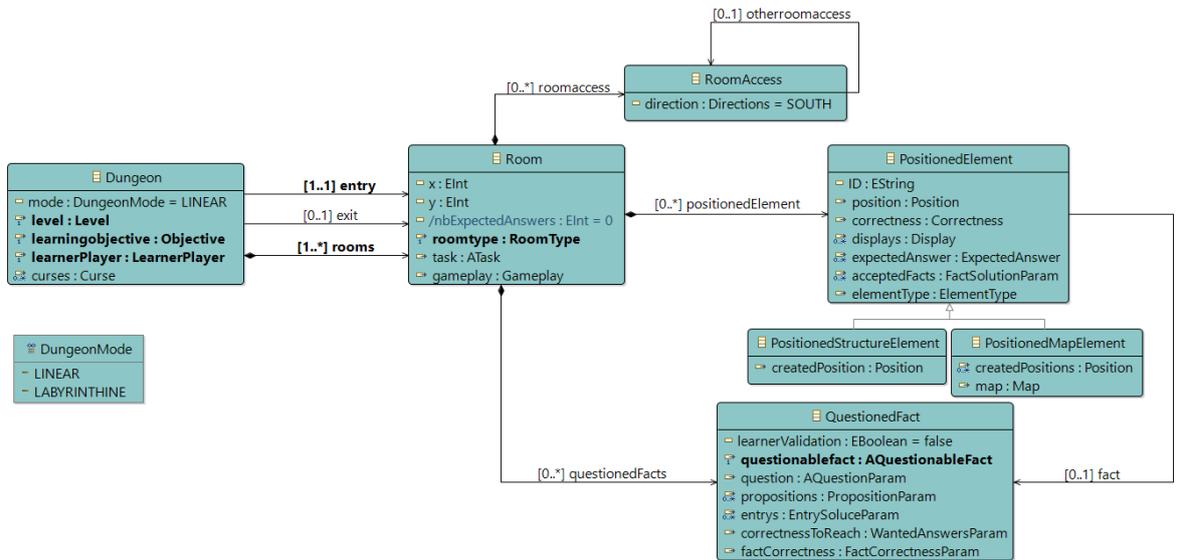


Figure 26 : Méta-modèle de l'activité

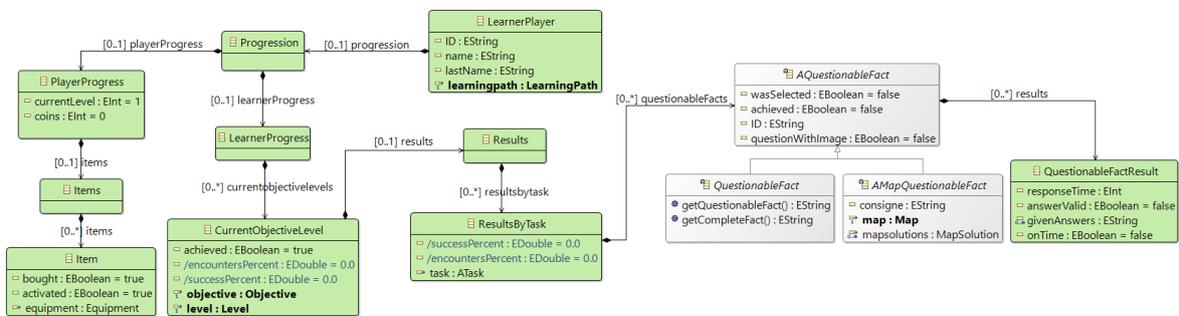


Figure 27 : Méta-modèle de l'apprenant-joueur

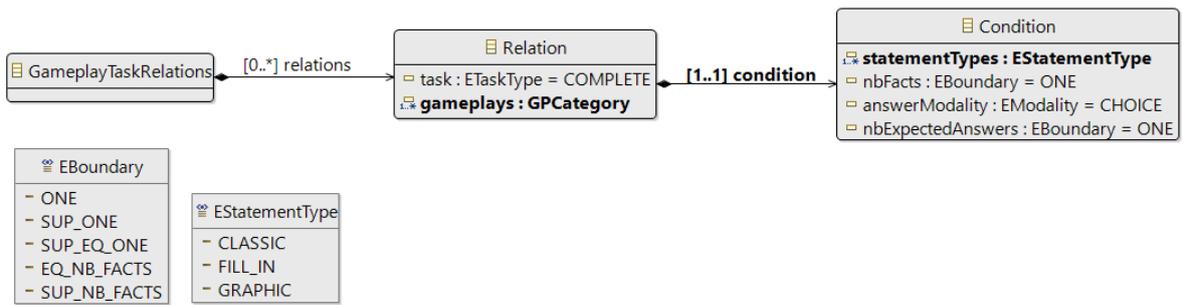


Figure 28 : Méta-modèle des relations

Modéliser et outiller la conception participative et générative de tableaux de bord d'apprentissage soutenant la création de sens

Modeling and Tooling Participatory and Generative Design of Learning Dashboards Supporting Sensemaking

Madjid SADALLAH¹ Jean-Marie GILLIOT¹

¹IIMT Atlantique, Lab-STICC UMR CNRS 6285, F-29238 Brest, France

Résumé. Les tableaux de bord d'apprentissage, dont le déploiement doit soutenir la prise de décision à toutes les étapes de la formation, continuent de faire face à des défis d'adoption. La littérature identifie plusieurs raisons à cette réticence, notamment l'absence de principes directeurs pour la conception et une participation insuffisante des parties prenantes. Pour y remédier, nous proposons un cadre de conception ainsi que deux outils pour supporter le processus de conception. Le cadre de conception, axé sur la création de sens pour étayer la prise de décision, guide la conception au travers d'un espace de conception exhaustif. Le premier outil propose une phase d'idéation participative et encourage la collaboration des différentes parties prenantes. Le second outil propose de simplifier le travail du développeur et d'encourager sa collaboration avec les parties prenantes. L'évaluation de ces outils repose sur des critères tels que la facilité d'utilisation, le soutien à la participation et l'expérience utilisateur, éclairant chaque étape du processus de conception. Les résultats montrent que ces outils soutiennent significativement la participation des utilisateurs finaux et l'expérience des développeurs, soulignant leur pertinence et leur efficacité. En favorisant l'adoption des tableaux de bord d'apprentissage par une conception centrée sur l'utilisateur, cette contribution met en lumière le potentiel de ces outils et l'importance des approches participatives dans le développement des technologies éducatives.

Mots-clés : tableaux de bord d'apprentissage, conception participative, conception générative, processus décisionnel, génération de tableaux de bord

Abstract. *Despite their potential, learning dashboards face challenges in terms of adoption. Research highlights a lack of design principles and insufficient stakeholder participation as key factors. To address these, we propose a design framework emphasizing sensemaking for decision-making. Additionally, we advocate for a collaborative design approach, supported by tools for ideation and prototyping. These tools aim to involve users actively, promoting an inclusive design approach. Their evaluation, based on ease of use, support for participation, and user experience, illuminates each design stage. Results show these tools significantly support end-user participation and developer experience, underscoring their relevance and effectiveness. By promoting the adoption of learning dashboards through user-centered design, this contribution not only unlocks their potential but also highlights the value of participatory approaches in educational technology development.*

Keywords: *learning analytics dashboards, participatory design, generative design, decision-making process, dashboard generation*

1. INTRODUCTION

Les tableaux de bord d'apprentissage (TBA), ou *Learning Analytics Dashboards*, sont des instruments conçus pour structurer et présenter des informations relatives aux apprenants, aux processus d'apprentissage et aux environnements éducatifs (Schwendimann *et al.*, 2017). Leur fonction première consiste à convertir ces données en informations visuelles, souvent sous forme graphique, afin de les rendre accessibles aux acteurs éducatifs. Cette transformation offre une vue précise et concise des performances des apprenants et des tendances d'apprentissage, ce qui facilite l'établissement d'objectifs clés et le suivi des progrès. Elle soutient également la prise de décisions éclairées et la mise en œuvre d'interventions ciblées (Jivet *et al.*, 2017). En fin de compte, ces outils aspirent à optimiser l'expérience d'apprentissage, les résultats des apprenants et à promouvoir une éducation plus éclairée et efficace grâce à l'utilisation stratégique des données.

L'intérêt pour les TBA a connu une croissance significative au fil des années. Cependant, malgré cette attention accrue, la recherche et l'exploration de ces outils demeurent limitées, entravant ainsi leur adoption à grande échelle (Alhamadi, 2020). Plusieurs facteurs contribuent à ces limitations. D'abord, en raison de leur émergence relativement récente, les TBA sont encore peu étudiés en termes de principes de conception (Echeverria *et al.*, 2018; Sedrakyan *et al.*, 2019). Or, au-delà de leur rôle d'agrégateurs de données, ces outils sont des systèmes complexes qui nécessitent une conception soignée pour garantir leur utilité et leur efficacité. L'absence de directives précises pour leur conception complique la création de TBA qui répondent aux exigences spécifiques des divers groupes d'utilisateurs, tels que les enseignants, les apprenants ou les administrateurs, chacun ayant des attentes et des profils uniques. Par ailleurs, un autre défi majeur réside dans la faible implication des utilisateurs en tant que parties prenantes dans le processus de conception des TBA (Alvarez *et al.*, 2020). En effet, les outils d'analyse de l'apprentissage peuvent imposer des hypothèses qui ne correspondent pas nécessairement aux besoins réels des utilisateurs (Gašević *et al.*, 2015), ces derniers possédant souvent une compréhension unique des préoccupations propres à leur contexte. Leur non-consultation ou intégration dans le processus de conception peut aboutir à des TBA qui ne répondent pas à leurs attentes, limitant ainsi leur adoption et leur efficacité. De plus, ces utilisateurs manquent souvent d'expérience dans l'utilisation de ces outils et de compétences en littératie des données (Schwendimann *et al.*, 2017). Enfin, un défi supplémentaire est la difficulté de développer des TBA capables de convertir les informations en actions facilement exploitables par les utilisateurs, en raison d'une compréhension limitée des processus décisionnels (Verbert *et al.*, 2020).

Pour relever ces défis, cet article se focalise sur la phase de conception des TBA, où l'implication précoce des utilisateurs finaux revêt une importance capitale. L'objectif est d'assurer une compréhension approfondie des besoins spécifiques de chaque catégorie d'utilisateurs, garantissant ainsi que les TBA répondent adéquatement à leurs attentes et sont conçus en tenant compte des réalités de leur contexte d'utilisation. Afin de renforcer l'autonomie des parties prenantes, nous proposons la mise en place d'un accompagnement ciblé visant à améliorer leurs compétences et leur compréhension de ces outils, favorisant ainsi une utilisation plus autonome et efficace. Dans ce contexte, la conception participative, également connue sous le terme de *codesign*, apparaît comme une approche prometteuse pour résoudre les défis liés à la conception et à l'utilisation des TBA (Prieto-Alvarez *et al.*, 2018). Cette approche met l'accent sur l'implication active de toutes les parties prenantes dans le processus de conception. En encourageant la communication et la collaboration entre les concepteurs, les enseignants, les apprenants et d'autres acteurs éducatifs, elle vise à établir un consensus. De plus, elle favorise le développement de modèles mentaux partagés, c'est-à-

dire des représentations internes des besoins, des attentes et des réalités propres au contexte d'utilisation des TBA, construites à partir des expériences passées, des connaissances préalables et des interactions sociales. Dans le cadre de la conception des TBA, cette construction collective de modèles mentaux facilite la compréhension commune des enjeux spécifiques à chaque utilisateur, ainsi que des solutions potentielles. Cela renforce la collaboration et la prise de décision en permettant à tous les acteurs de contribuer à la conception des TBA de manière informée et réfléchie. De plus, la conception participative vise à garantir une utilisation réfléchie des données générées par les TBA. Elle encourage les utilisateurs à interpréter les informations de manière critique et à les aligner sur des objectifs éducatifs spécifiques. Ainsi, cette approche favorise une utilisation plus autonome et efficace des TBA, tout en répondant aux besoins diversifiés des parties prenantes dans le domaine de l'éducation.

Malgré la reconnaissance croissante de l'importance des méthodes de conception participative parmi les chercheurs et praticiens, et quelques exemples éducatifs soulignant leur utilité (Sarmiento et Wise, 2022), leur intégration dans le domaine de l'analyse de l'apprentissage n'est pas encore largement répandue (2022). Les obstacles à cette intégration sont multiples, notamment une compréhension peu claire des approches spécifiques à adopter (Alvarez *et al.*, 2020; Dollinger *et al.*, 2019). Plus particulièrement, dans le contexte des TBA, l'adoption limitée de la conception participative s'explique en partie par le manque de clarté sur la manière d'appliquer efficacement cette approche dans ce domaine (Alvarez *et al.*, 2020). En outre, l'absence d'outils spécifiquement conçus pour répondre aux besoins des acteurs de l'éducation constitue un autre problème majeur (Dollinger *et al.*, 2019). Ces obstacles freinent la diffusion et l'adoption à grande échelle de ces méthodes, malgré leur potentiel considérable à améliorer la conception et l'efficacité des TBA.

Dans ce contexte, notre objectif est de combler ces lacunes en proposant des outils spécifiques pour faciliter la conception participative des TBA. Cette approche, qui favorise une collaboration étroite entre les concepteurs et les utilisateurs finaux, vise à aligner les fonctionnalités du TBA sur les besoins et les attentes des acteurs de l'éducation. En adoptant cette approche, nous aspirons à promouvoir une utilisation plus efficace des TBA en éducation. Cette efficacité visée reflète la capacité des TBA co-conçus à répondre aux besoins des utilisateurs finaux dans le domaine de l'apprentissage, se matérialisant à plusieurs niveaux. Tout d'abord, en termes de qualité de l'information, les TBA doivent fournir des données pertinentes et exploitables pour permettre aux utilisateurs de comprendre les tendances, d'identifier les points forts et les faiblesses, et de prendre des décisions éclairées. Ensuite, en termes d'utilisabilité, ils doivent être conviviaux et accessibles, même pour ceux qui ne sont pas familiers avec de tels outils, afin de favoriser leur adoption et leur utilisation régulière. Enfin, en termes d'impact, les TBA doivent avoir un effet mesurable sur les résultats de l'apprentissage, contribuant ainsi à l'amélioration des performances des étudiants, à la performance des programmes d'enseignement, ou à d'autres objectifs spécifiques définis dans le contexte de l'éducation. Pour soutenir cette efficacité, nous nous concentrons sur deux aspects principaux : premièrement, la mise en place d'une approche participative dès les phases de conception, favorisant ainsi une meilleure adaptation des outils aux besoins réels des acteurs éducatifs ; deuxièmement, le développement d'outils spécifiques qui soutiennent cette démarche participative, permettant une collaboration efficace entre les concepteurs, les développeurs et les parties prenantes. En tenant compte de ces aspects, notre recherche aspire à surmonter les obstacles identifiés et à ouvrir la voie à une utilisation plus répandue et plus efficace des TBA.

Cette recherche a adopté une méthodologie de *Design-Based Research* (DBR), couramment utilisée pour étudier les environnements éducatifs à travers des cycles itératifs de conception, de mise en œuvre et d'analyse (Wang et Hannafin, 2005). La première itéra-

tion, rapportée dans (Gilliot *et al.*, 2018), 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 (Shaw, 2012), jouant le rôle de modèle commun. L'itération actuelle vise à atteindre plusieurs objectifs : (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 entre les parties prenantes ; (3) développer un outil de prototypage opérationnel LADStudio ; et (4) analyser les outils développés.

La structure de cet article se décline comme suit. La section 2 se concentre sur la conception participative dans le domaine de l'éducation, mettant en évidence ses caractéristiques distinctives et son utilisation spécifique dans le cadre des TBA. Nous y détaillons notre méthodologie et notre modèle d'interaction visant à soutenir le processus décisionnel à travers les TBA, ainsi que notre cadre de conception dédié. La section 3 est dédiée aux outils que nous avons élaborés pour faciliter le processus participatif, en abordant les phases d'idéation et de prototypage. Dans la section 4, nous détaillons les expérimentations réalisées pour évaluer notre approche et nos outils, suivies d'une analyse détaillée des résultats. Enfin, la section 5 présente nos conclusions et propose des directions pour les recherches futures.

2. CONCEPTION PARTICIPATIVE DE TABLEAUX DE BORD D'APPRENTISSAGE SUPPORTANT LA DÉCISION

Dans cette section, nous définissons d'abord la conception participative, soulignons ses caractéristiques distinctives et explorons ensuite son application dans le domaine de l'éducation en analysant ses méthodes et ses outils. Ensuite, nous exposons notre méthodologie, conçue pour impliquer activement les parties prenantes en leur permettant d'exprimer leurs besoins et de concrétiser leurs idées à travers des prototypes. Cette méthodologie s'appuie sur les principes fondamentaux établis dans le domaine de la conception participative en éducation, tout en intégrant des aspects cruciaux relatifs à l'utilisation des TBA lors du processus de conception. Nous accordons notamment une attention particulière à la *création de sens* (ou *sensemaking*) et au soutien à la prise de décision à travers des visualisations pertinentes, éléments indispensables pour optimiser l'impact des TBA lors de leur utilisation (Meyer *et al.*, 2010). Enfin, nous détaillons les composants essentiels d'un cadre de conception. Ce cadre, en délimitant clairement l'espace de conception, vise à structurer les différentes étapes du processus de conception participative, marquant ainsi la transition vers nos propositions concrètes pour une approche plus efficace de la conception des TBA.

2.1. CONCEPTION PARTICIPATIVE EN ÉDUCATION

La *conception participative*, qui découle de la conception centrée sur l'utilisateur, englobe les pratiques de conception qui impliquent activement les parties prenantes tout au long du processus, depuis l'exploration initiale jusqu'au développement final. Cette approche promeut un apprentissage mutuel entre les participants et encourage leur engagement dans des processus collectifs de conception (Robertson & Simonsen, 2012, p. 2). La transition vers une telle approche marque un changement de paradigme, passant de la conception pour les utilisateurs à la conception avec les individus comme partenaires égaux (Sanders et Stappers, 2008). Cette évolution vise à intégrer les participants en les reconnaissant comme des contributeurs légitimes au sein de l'équipe de conception (Visser *et al.*, 2005). En tant qu'approche de *co-création*, les concepteurs, formés à la créativité, collaborent avec des non-concepteurs tout au long du processus de conception. L'accent est donc mis sur l'apprentissage mutuel : les utilisateurs découvrent les possibilités et les contraintes de la technologie, tandis que les

concepteurs acquièrent une compréhension des pratiques et des perspectives des utilisateurs (Schuler et Namioka, 1993). La recherche a également montré que l'implication des utilisateurs finaux dans le cycle de conception assure des solutions de conception adaptées, alignées sur leurs besoins et attentes (Holstein *et al.*, 2017 ; Taffe, 2015).

La conception participative en éducation, qui se caractérise par son parcours historique propre, sa philosophie intrinsèque et sa méthodologie (DiSalvo *et al.*, 2017), représente une tendance en pleine expansion. Cette approche met l'accent sur l'importance d'une collaboration active entre tous les acteurs de l'éducation. Elle est articulée autour de plusieurs étapes clés, telles que l'idéation, le prototypage et le test, nécessitant chacune des outils et des méthodes adaptés pour guider efficacement le processus. Parmi les techniques populaires utilisées pour la conception participative en éducation, on retrouve notamment les ateliers et les groupes de discussion, les personas d'apprentissage, les outils basés sur des cartes, ainsi que les croquis et les prototypes.

Les *ateliers et les groupes de discussion* sont couramment utilisés pour générer des idées de conception et recueillir les avis des participants. Ces méthodes permettent une interaction directe avec les utilisateurs finaux, favorisant ainsi une compréhension approfondie de leurs besoins et attentes. Par exemple, pré-interviewer les utilisateurs finaux sur leurs besoins peut être utilisé pour dériver des idées de conception (Xhakaj *et al.*, 2016).

Le *persona d'apprentissage* permet de modéliser et de résumer les informations essentielles sur les utilisateurs finaux potentiels. Lors de sessions de codesign, les participants peuvent être engagés à travailler collectivement pour créer de nouveaux profils de persona ou pour enrichir des instances existantes (Prieto-Alvarez *et al.*, 2018). Ces représentations peuvent ensuite être utilisées pour guider la conception et le développement de nouvelles fonctionnalités ou améliorations.

Le *codesign basé sur des cartes* est une autre approche populaire. Les cartes assurent une base commune pour la compréhension et la communication au sein d'une équipe, fournissant ainsi des résumés pratiques d'informations et/ou de méthodes utiles. L'utilisation de cartes peut renforcer la collaboration, stimuler la créativité combinatoire et faciliter l'exploration des liens possibles entre les besoins de décision des utilisateurs et les nombreuses propositions d'indicateurs et de visualisations disponibles dans les ateliers de conception (Lucero *et al.*, 2016). Par exemple, LA-DECK (Alvarez *et al.*, 2020) est un outil de co-conception basé sur des cartes, conçu spécifiquement pour la création d'outils d'analyse de l'apprentissage.

Les *croquis et les prototypes* sont d'autres techniques utilisées pour concrétiser les résultats des efforts collaboratifs entre participants. Ce processus de réalisation permet de créer des représentations visuelles et d'impliquer les participants au-delà de la phase de génération d'idées. Il leur permet de communiquer de manière non conventionnelle, en les incitant à matérialiser collectivement leurs attentes (Gaver *et al.*, 1999).

Notre méthodologie, détaillée dans la suite de cet article, se fonde sur ces principes théoriques et ces méthodes pratiques pour guider notre démarche de conception participative des TBA. L'objectif est de transcender la simple fonction d'outil d'analyse des TBA pour en faire des ressources qui favorisent une compréhension approfondie des données et une prise de décision éclairée. En intégrant ces concepts, notre méthodologie vise à orienter la conception de TBA qui ne se limitent pas à présenter des données, mais qui les rendent accessibles, compréhensibles et utiles pour les utilisateurs, contribuant ainsi à une amélioration tangible en éducation.

2.2. MÉTHODOLOGIE DE CONCEPTION PARTICIPATIVE DES TBA

En matière de conception des TBA, l'approche participative permet d'aligner les différents acteurs éducatifs sur les objectifs principaux du tableau de bord, facilite une mise en

œuvre cohérente et anticipe les éventuels obstacles à son adoption (Boscardin *et al.*, 2018). Pour enrichir notre processus de conception des TBA, nous avons intégré de manière synergique une approche de conception générative à notre démarche participative. Cette combinaison méthodologique vise à dynamiser la création des TBA en favorisant une collaboration étroite avec les parties prenantes dès le début du processus.

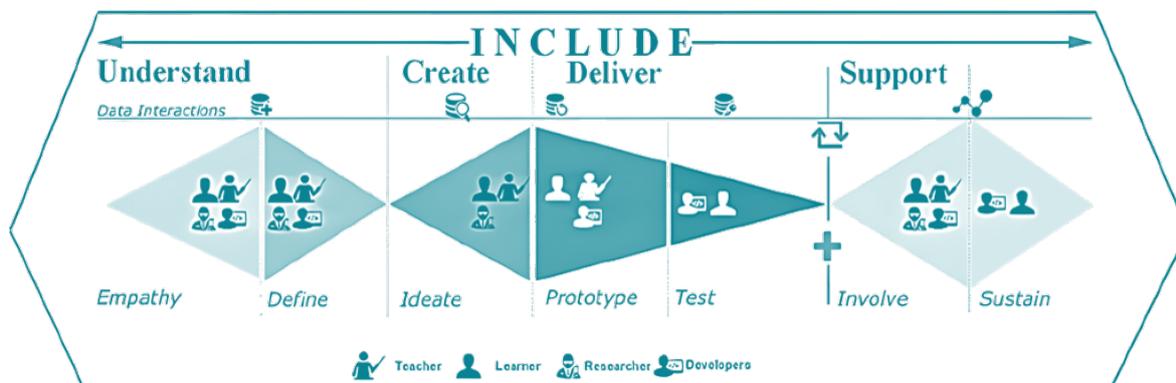


Figure 1 : Processus de co-design en analytique de l'apprentissage (2018)

La *conception participative* constitue la première pierre angulaire de notre démarche. Elle vise à impliquer activement les parties prenantes dès le début, recueillant leurs idées et besoins pour une compréhension approfondie tout en favorisant leur engagement continu. Nos propositions s'inscrivent dans le cadre du *Modèle de codesign en analytique de l'apprentissage*, élaboré par Prieto-Alvarez *et al.* (2018). Ce modèle décompose le processus de conception participative en une série d'activités rassemblant les différents acteurs impliqués, visant à affiner continuellement les exigences et à se rapprocher progressivement de la solution recherchée. Notre application spécifique de ce modèle se concentre sur les TBA. La phase préliminaire de compréhension (*Understand*), initiée à travers des entretiens d'exploration (Gilliot *et al.*, 2018), oriente nos phases ultérieures d'idéation et de prototypage.

La *conception générative* complète cette première phase en fournissant aux parties prenantes des outils puissants pour articuler de manière approfondie leurs besoins et intentions. Elle permet d'explorer un espace de conception étendu, facilitant la génération de multiples solutions cibles (Robertson et Simonsen, 2012). En associant ces deux approches, nous établissons un cadre méthodologique solide, favorisant la collaboration et la génération itérative de solutions innovantes, alignées sur les besoins spécifiques des parties prenantes. L'intégration de la *conception générative* apporte une dimension créative et exploratoire au processus, permettant aux parties prenantes de participer activement à la création et à la sélection des solutions, renforçant ainsi leur engagement dans le projet (Sanders et Stappers, 2008).

2.3. CRÉATION DE SENS AU CENTRE DE LA CONCEPTION DES TBA

Envisager les TBA comme des instruments de communication requiert une exploration approfondie de la cognition humaine, de l'interaction homme-machine et des technologies de visualisation (Alhadad, 2018 ; Yoo *et al.*, 2015). En effet, pour maximiser leur impact, les TBA doivent être conçus sur des bases théoriques solides (Alhadad, 2018 ; Yoo *et al.*, 2015), afin de renforcer la conscience et la réflexion des utilisateurs, et d'entraîner des modifications dans leurs compétences cognitives, comportementales et émotionnelles (Jivet *et al.*, 2017). Par conséquent, une exploration approfondie des principes de conception des TBA est cruciale pour guider ces choix (Bodily et Verbert, 2017 ; Echeverria *et al.*, 2018). L'efficacité de la conception des TBA pour une communication réussie dépend d'une compréhension

approfondie de la manière dont les humains assimilent l'information (Few, 2006). Examiner comment les TBA peuvent favoriser des aspects essentiels tels que la création de sens et la prise de décision, en mettant l'accent sur l'interaction de l'utilisateur, est par conséquent essentiel pour toute approche de conception.

2.3.1. Création de sens et théorie Data/Frame

La création de sens, ou *sensemaking*, est un processus méta-cognitif fondamental dans l'utilisation des TBA. C'est un processus dynamique et réciproque qui implique l'adaptation des données à un cadre cognitif existant (ou modèle mental) et la modification de ce cadre en réponse aux nouvelles données (Pirolli et Card, 2005). En d'autres termes, il s'agit d'un processus d'interprétation des données à travers le prisme d'un modèle mental existant, tout en révisant ce modèle en fonction des nouvelles informations que les données apportent. Les modèles existants qui explorent la stimulation de la pensée humaine à travers des outils de visualisation décomposent le processus de création de sens en différentes phases, allant de la perception initiale de l'information à la prise de décision et à la mise en œuvre d'actions pédagogiques (Campos *et al.*, 2021 ; Wise et Jung, 2019). Pour approfondir notre analyse de la création de sens liée aux TBA, nous nous appuyons sur la théorie *Data/Frame* (D/F) proposée par Klein *et al.* (2006). Cette théorie offre un éclairage précieux sur les mécanismes cognitifs en jeu, en articulant la dynamique entre les données — perçues comme des signaux ou des indicateurs d'événements — et les cadres cognitifs — envisagés comme des structures interprétatives qui donnent du sens aux données et orientent la quête de nouvelles informations (Ndukwe et Daniel, 2020).

La pertinence de la théorie D/F pour l'analyse de la création de sens via les TBA est à la fois multiple et significative. Tout d'abord, cette théorie décrit de manière précise la façon dont les données, sous forme de signaux, interagissent avec les cadres cognitifs, qui sont les mécanismes par lesquels les individus interprètent et attribuent un sens aux informations (Klein *et al.*, 2006). Cette interaction constitue le cœur de l'utilisation des TBA, où les données doivent être interprétées de manière à éclairer la prise de décision. Deuxièmement, la théorie D/F reconnaît que le processus de *sensemaking* est une activité intentionnelle, menée par des agents poursuivant des objectifs spécifiques (Pirolli et Card, 2005). Cette caractéristique est fondamentale pour les utilisateurs de TBA, qui déploient ces outils dans des buts précis tels que l'amélioration de l'enseignement ou de l'apprentissage. Troisièmement, la théorie D/F aborde la question des biais cognitifs, ces pièges de raisonnement qui peuvent influencer négativement l'interprétation des données (Klein *et al.*, 2006). Comprendre ces biais est crucial pour appréhender pourquoi certains utilisateurs peuvent mal interpréter les informations fournies par les TBA. Enfin, par rapport à d'autres modèles de *sensemaking*, la théorie D/F se distingue par son approche pragmatique et représentationnelle, se concentrant sur la façon dont les individus construisent du sens à partir de leur expérience face à des problématiques concrètes. Cette caractéristique en fait un cadre conceptuel particulièrement adapté à l'étude de la création de sens dans le contexte spécifique des TBA.

2.3.2. Conscience de la situation et interaction utilisateur

La création de sens, un processus essentiel dans diverses activités humaines, est particulièrement cruciale lorsqu'il s'agit de déchiffrer des situations complexes pour prendre des décisions éclairées et appropriées. Cette capacité joue un rôle important dans l'utilisation des outils d'analyse en éducation. La conscience de la situation (*Situational Awareness* - SA) est un concept crucial dans la prise de décision, notamment dans des situations où l'information doit être traitée rapidement et avec précision. Développée par Endsley (1995), la SA se com-

pose de trois niveaux interdépendants qui facilitent la compréhension et l'anticipation des événements, permettant ainsi une prise de décision efficace :

1. *Perception des éléments dans l'environnement* : cette étape initiale implique la détection des signaux et des indices dans l'environnement immédiat, ce qui est crucial pour toute analyse ultérieure.
2. *Compréhension de la signification de ces éléments* : au-delà de la perception, cette phase englobe l'interprétation et l'intégration des informations perçues, en établissant des liens entre les éléments et en évaluant leur pertinence par rapport aux objectifs poursuivis.
3. *Projection de l'état futur des éléments perçus* : cette dernière phase consiste à anticiper les événements futurs en se basant sur la compréhension actuelle, ce qui est essentiel pour la planification stratégique et la préparation aux actions nécessaires.

La théorie de la création de sens de Klein *et al.* (2006) suggère que ces niveaux de SA ne deviennent pleinement exploitables que lorsque les individus activent leurs cadres cognitifs pour donner un sens aux informations perçues. Cela implique souvent de remettre en question les hypothèses existantes, de reconnaître les modèles émergents et d'ajuster les cadres cognitifs en fonction des nouvelles données.

2.3.3. Modèle d'interaction pour la création de sens soutenant le processus décisionnel par les TBA

L'exploration du processus de création de sens via les TBA nécessite une analyse détaillée des interactions entre l'utilisateur et l'outil. Les TBA, qui sont conçus pour organiser les données afin de faciliter leur interprétation, offrent la possibilité de tirer des conclusions et de planifier des actions ayant un impact sur l'enseignement et l'apprentissage. Toutefois, les modèles actuels ne reflètent pas entièrement l'intégration des activités cognitives dans l'expérience utilisateur des TBA. Or, comprendre comment les utilisateurs interagissent avec ces outils est crucial, car la création de sens implique la recherche de représentations pertinentes et l'encodage des données pour répondre à des questions spécifiques à la tâche (Russell *et al.*, 1993). Enfin, une compréhension précise de ces interactions est nécessaire pour optimiser l'efficacité des TBA en éducation, afin d'ajuster les modèles existants en tenant compte de la dynamique complexe entre les activités cognitives, la recherche de représentation et les interactions utilisateur-TBA.

Les interactions avec les TBA jouent un rôle essentiel en permettant aux individus d'explorer les représentations visuelles, générant ainsi des connaissances et des compréhensions cruciales pour le processus décisionnel. En établissant une connexion étroite entre l'analyse, la visualisation et l'analyste humain, ces interactions favorisent une approche intégrée de la prise de décision (Endert *et al.*, 2014). Pour examiner comment les utilisateurs participent au processus de création de sens avec les TBA, nous adoptons une perspective qui va au-delà des aspects techniques et modaux de l'interaction. Nous envisageons l'interaction comme un échange dynamique entre les utilisateurs et les TBA, un dialogue continu qui façonne le processus de création de sens, en accord avec la théorie de la *cognition distribuée* (Hutchins, 1995). Selon cette théorie, la cognition émerge de l'interaction entre les individus et leur environnement, se développant au fil du temps par la perception et l'action (Liu *et al.*, 2008). Ainsi, les utilisateurs et les TBA forment un système interactif où la coordination et l'influence causale se manifestent à travers l'échange de représentations, offrant une perspective intéressante sur les processus cognitifs mobilisés.

En envisageant les TBA comme des outils cognitifs, nous proposons un modèle de création de sens qui s'appuie sur les TBA, comme illustré à la Figure 2. Dans ce modèle, les

TBA jouent le rôle d'instruments cognitifs qui aident les utilisateurs à détecter visuellement des motifs, des tendances et des anomalies dans les données, les orientant ainsi vers des décisions informées et pertinentes. La création de sens est perçue comme un processus cognitif qui fournit à la fois un moyen d'interpréter les données environnementales et un catalyseur pour agir en conséquence. L'utilisateur interagit avec le tableau de bord pour recueillir des informations et effectuer des actions basées sur les données, déclenchant ainsi un processus actif de création de sens. Ce processus peut être décomposé en deux tâches imbriquées :

1. *Interaction* : durant cette phase, l'utilisateur interagit avec le TBA pour rechercher, filtrer, lire et extraire des informations des représentations affichées. Cette interaction initie des processus en arrière-plan qui recalculent la mise en page graphique du TBA pour ajuster son affichage.
2. *Cadrage (framing)* : l'utilisateur intègre les données collectées lors des interactions à ses connaissances internes pour construire, modifier ou consolider un cadre mental. La construction du cadre, un processus itératif, évolue en fonction des données obtenues, conduisant potentiellement à de nouvelles interactions avec le TBA.

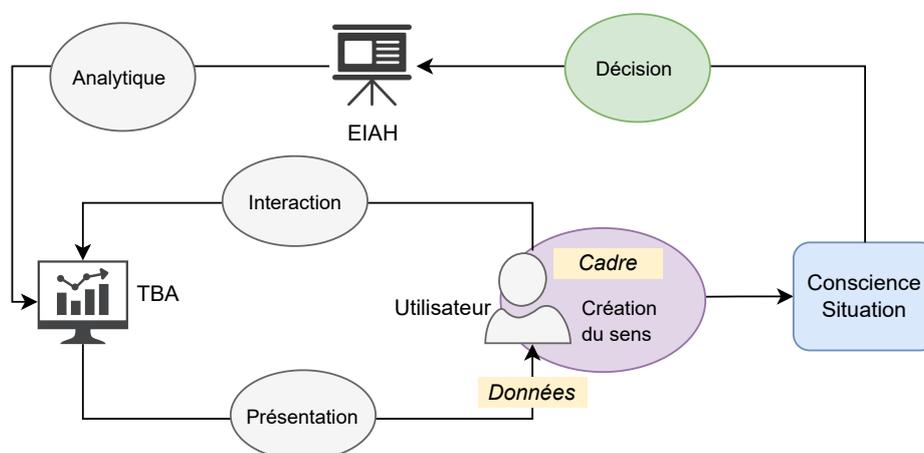


Figure 2 : Modèle d'interaction pour la création de sens pris en charge par les TBA

Pour décrire la construction de cadres mentaux (*framing*), nous nous appuyons sur les activités de création de sens identifiées par la théorie D/F. Ces activités incluent l'intégration de nouvelles données pour enrichir la compréhension existante, l'examen critique des informations en désaccord avec le cadre mental préétabli, le maintien d'une interprétation cohérente malgré les contradictions, la considération de diverses interprétations possibles pour déterminer la plus convaincante, l'ajustement ou la révision du cadre mental pour intégrer des informations contradictoires, et le développement d'une compréhension renouvelée pour résoudre les tensions entre les différentes données. Ces processus cognitifs reflètent la nature dynamique et itérative de la création de sens, se déroulant à travers une interaction continue entre l'utilisateur et le TBA. Ils soulignent également l'importance de l'adaptabilité cognitive, essentielle pour ajuster les cadrages mentaux en réponse aux nouvelles informations et aux défis analytiques posés par les données. Cette flexibilité se révèle indispensable pour naviguer dans l'environnement complexe et souvent ambigu des données analytiques.

2.4. CADRE DE CONCEPTION DE TBA

Cette section présente DEFLAD (Gilliot et Sadallah, 2024), un cadre de conception visant à clarifier les dimensions de conception des TBA, tout en intégrant des capacités de

création de sens et en mettant en lumière celles liées aux processus décisionnels. Le modèle d'interaction intégré active la création de sens, renforçant ainsi le processus décisionnel inhérent aux TBA, les considérant comme des outils cognitifs essentiels. Établi à travers un examen approfondi des dimensions de conception des TBA, ce cadre permet de définir un espace de conception complet qui servira de base aux outils proposés.

2.4.1. Espace de conception de TBA

Le concept d'espace de conception trouve ses racines dans la *Théorie du Traitement de l'Information* développée par Newell et Simon (1972). Selon cette théorie, la résolution d'un problème est envisagée comme une exploration dans un espace d'états clairement défini associé à ce problème. Dans le contexte de la conception, cette exploration devient une démarche essentielle. Pour un problème de conception donné, l'espace de conception englobe toutes les solutions envisageables pour l'artefact à concevoir, y compris les différentes alternatives possibles pour chaque solution (Shaw, 2012).

Une approche courante pour définir un espace de conception consiste à extraire des principes de conception communs des techniques existantes dans le domaine ciblé (Schulz *et al.*, 2011). En matière de conception de tableaux de bord, de nombreux chercheurs ont entrepris de caractériser cet espace. Par exemple, dans le domaine de la visualisation des données, Schulz *et al.* (2011) ont élaboré un modèle d'espace de conception en décrivant comment les concepteurs peuvent raisonner sur les tâches à travers six dimensions d'analyse des tâches : le « Pourquoi » (objectif de la tâche), le « Comment » (moyens de la tâche), le « Quoi » (caractéristiques des données), le « Où » (la cible et la cardinalité des entités de données dans cette cible), le « Quand » (ordre des tâches) et le « Qui » (l'utilisateur ou le type d'utilisateur). Ces dimensions sont également abordées dans les travaux de Sarikaya *et al.* (2019), qui introduisent également la littératie en matière de visualisation des utilisateurs et proposent une analyse approfondie des interactions. L'espace de conception défini par ces auteurs comprend quatre dimensions de conception liées à différents aspects d'un tableau de bord : l'objectif (décisions stratégiques, tactiques et opérationnelles, et apprentissage), le public (circulation, littératie en matière de visualisation requise et expertise requise dans le domaine), les fonctionnalités visuelles (construction et composition, pages multiples, interface interactive, mise en évidence et annotation, modification de l'état des données ou du monde) et la sémantique des données (alertes et notifications, points de repère, mise à jour).

Certains auteurs suggèrent d'adopter une approche basée sur les questions des 5 W – Who (*Qui*)?, When (*Quand*)?, Why (*Pourquoi*)?, What (*Quoi*)?, hoW (*Comment*)? – pour explorer l'espace de conception (Chatti *et al.*, 2020). Cette approche se concentre sur l'objet d'intérêt, allant de la visualisation aux tableaux de bord et à l'analyse des processus d'apprentissage. Dans DEFLAD, nous formulons les 5 W de la manière suivante :

1. *Qui* ? englobe le public cible et les interactions entre différents utilisateurs ;
2. *Quand* ? permet de déterminer si l'utilisation se fait en temps réel ou est différée ;
3. *Pourquoi* ? correspond à l'objectif spécifique du TBA ;
4. *Quoi* ? détaille le contexte d'utilisation du TBA ainsi que les données pertinentes ;
5. *Comment* ? se concentre sur la visualisation, étant étroitement lié à la création de sens à travers l'utilisation du TBA.

DEFLAD offre une approche intégrée pour la conception de TBA qui soutient efficacement les tâches de création de sens. Il permet également de définir clairement l'espace de conception, assurant ainsi une approche structurée et exhaustive pour la conception de TBA. Comme illustré sur la Figure 3, DEFLAD établit un espace de conception en deux parties

distinctes. La première partie, qui comprend les trois premières questions, se concentre sur la définition du contexte d'utilisation de TBA. La seconde partie, qui englobe les deux dernières questions, traite des aspects de visualisation et d'interaction. Le Tableau 1 synthétise les propriétés que nous avons identifiées pour chacune des dimensions de conception. Les paragraphes qui suivent procèdent à une analyse approfondie et à une discussion détaillée de chaque dimension.

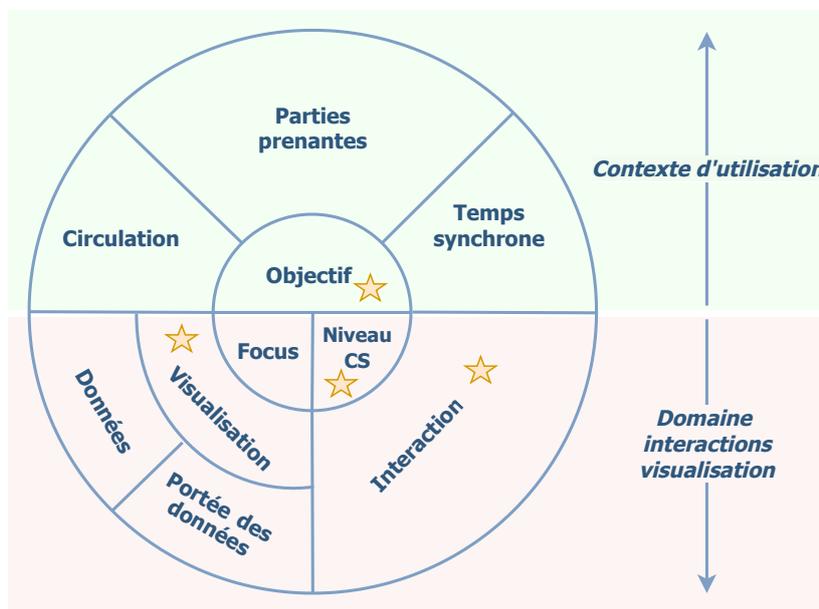


Figure 3 : Cadre de conception de tableaux de bord d'apprentissage DEFLAD. Les étoiles représentent les options de soutien à la création du sens

2.4.2. Qui ? Identification du public

Deux dimensions sont liées au public : (1) l'identification des utilisateurs de TBA, qui est classique en analyse de l'apprentissage ; et (2) la circulation des TBA, qui reconnaît leur potentiel en tant qu'outils de communication.

UTILISATEURS. Les TBA peuvent cibler divers acteurs, tels que les administrateurs, les instructeurs, les apprenants, voire tous ces groupes. Des études antérieures ont défini quatre catégories d'acteurs liées aux TBA : le niveau méga (gouvernance), le niveau macro (institutionnel), le niveau méso (curriculum et enseignant/tuteur) et le niveau micro (apprenant) (Ifenthaler et Widanapathirana, 2014 ; Sedrakyan *et al.*, 2019). En reconnaissant que les TBA peuvent servir d'outil de communication entre ces différents acteurs, le concept de *public cible* élargit la notion traditionnelle d'utilisateur (Sarikaya *et al.*, 2019).

CIRCULATION. En relation avec le concept de « public », la circulation interpersonnelle d'un tableau de bord peut être subdivisée en quatre catégories : *Public*, *Organisationnel*, *Collectif*, et *Individuel*. Chaque catégorie devient plus spécifique et nécessite un contexte supplémentaire, qui peut ne pas être inclus dans le tableau de bord lui-même (2019). Un tableau de bord *Public* est destiné à une présentation générale et peut présenter des données d'importance sociétale. Les tableaux de bord de type *Organisationnel* sont pertinents dans le cadre d'une structure organisationnelle pour des personnes ayant un objectif commun (par exemple, la validation des diplômes). La catégorie *Collectif* englobe les contextes où

un individu limite l'accès au tableau de bord aux personnes de son choix, identifiant des scénarios de données ou d'analyses sensibles. Enfin, la catégorie *Individuel* comprend les tableaux de bord qui quantifient l'individu et ne sont généralement pas partagés, sauf avec des personnes de confiance, comme un enseignant dans le cas d'un apprenant.

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

Dimension	Éléments	Valeurs
Qui ?	Utilisateur	Gouvernance, Institution, Curriculum, Enseignant/tuteur, Apprenant
	Circulation	Public, Organisationnel, Collectif, Individuel
Quand ?	Temps réel	O/N
Pourquoi ?	Focus	Processus d'apprentissage (cognitif, axé sur les résultats, axé sur le processus, comportemental, méta-cognitif, social, émotionnel) Gestion (personnes, ressources, activités, expérience)
	Niveau de SA	Perception (ou surveillance, prise de conscience, soutien), Compréhension (ou analyse, réflexion), Action (ou projection, décision, intervention, rétroaction, évaluation)
Quoi ?	Données	Liste des données pertinentes
	Portée des données	Apprenant, Enseignant, Salle de classe, Institution
	Source de données	Salle de classe, Systèmes de gestion de l'apprentissage, Curriculum, Profil, Autre
	Durée des données	Une session, Un semestre, Une année, Scolarité complète, Tout au long de la vie
Comment ?	Visualisation	Type de diagramme
	Interaction	Zoom, Filtrer, Détails à la demande, Relier, Historique, Extraire

2.4.3. Quand ?

Cet aspect de l'espace de conception fait référence au moment d'utilisation du tableau de bord. En lien avec le public cible, il s'agit du point de rencontre entre les utilisateurs et les données. On peut déterminer si la communication est basée sur le traitement des données en temps réel, c'est-à-dire si elle se concentre sur les événements actuels, ou si elle repose sur des données historiques.

2.4.4. Pourquoi ? Raffinement des objectifs des TBA

Cette dimension est intrinsèquement liée à l'objectif du TBA. Bien que cet aspect soit central dans les TBA et fournisse la justification de leur conception, il demeure ambigu dans

la littérature existante. Certains auteurs parlent de *finalité*, d'autres d'*objectifs*, voire de *but*. Nous choisissons de conserver le terme « objectif ».

Nous soutenons que les objectifs des TBA se réfèrent à deux dimensions de l'espace de conception : le *Focus* et le *Niveau de conscience de la situation* (SA). Ces deux dimensions permettent d'intégrer toutes les options existantes et d'envisager de nouvelles. Concrètement, elles (i) ouvrent de nouvelles opportunités de conception et (ii) encouragent les concepteurs à aller au-delà d'options simples comme la *surveillance*, et à explorer des niveaux plus élevés de conscience de la situation.

FOCUS. En termes d'objectifs des TBA, la plupart des travaux se limitent à fournir des listes d'exemples, généralement dérivés des travaux de Park et Jo (2015). Sedrakyan *et al.* (2019) ont analysé les objectifs des étudiants sous l'angle de l'intervention visée (ou de la rétroaction) et ont proposé les points de focus suivants en lien avec le processus d'apprentissage : (i) cognitif, (ii) orienté vers les résultats (par exemple, le niveau de réussite), (iii) orienté vers le processus, (iv) comportemental, (v) méta-cognitif. Comme la présence sociale est un élément clé de l'expérience éducative (Garrison *et al.*, 2003), nous proposons d'ajouter à cette liste (vi) l'aspect social, qui concerne le travail de groupe ou les relations entre apprenants (par exemple, dans les forums). Les interventions sur (vii) les émotions (Ez-Zaouia *et al.*, 2020) devraient compléter cette liste. Cette extension de la liste des objectifs des TBA vise à mieux prendre en compte les aspects sociaux et émotionnels de l'expérience d'apprentissage des étudiants, contribuant ainsi à une conception plus holistique et centrée sur l'utilisateur de ces outils.

Les objectifs des enseignants incluent des interventions liées au processus d'apprentissage des étudiants ; cette même liste de points de focus s'applique également aux étudiants. Ils sont également préoccupés par la gestion des cours, et leurs objectifs peuvent se concentrer sur (i) les personnes (par exemple, les étudiants à risque), (ii) les ressources (gestion ou amélioration), (iii) les activités (y compris l'évaluation), (iv) l'expérience d'apprentissage vécue. D'autres parties prenantes au niveau méga- ou macro-niveau peuvent s'intéresser à des objets de focus supplémentaires, mais cette liste reste à proposer.

NIVEAU DE CONSCIENCE DE LA SITUATION. Généralement, un objectif peut être exprimé sous la forme d'un verbe suivi d'un objet. Lors de l'examen des listes d'objectifs, l'expression peut être liée à un niveau de conscience de la situation ou d'état (Endsley, 1995), qui peut prendre différents noms en fonction du contexte : (1) *sensibilisation* ou surveillance, (2) *analyse* ou réflexion, et (3) *action*, ou projection, décision, intervention, impact, rétroaction, évaluation.

Le Tableau 2 présente la liste des objectifs élaborés par Park et Jo (2015) en fonction de leur *Focus* et du *Niveau de conscience de la situation* visé. Il couvre la plupart des options de processus d'apprentissage, mais pas les niveaux de conscience de la situation correspondants. Les utilisateurs devraient être incités à explorer d'autres niveaux de conscience de la situation, en tenant compte de la rétroaction. En ce qui concerne les options de gestion, il existe encore de nombreuses possibilités à étudier.

2.4.5. Quoi ? Spécification des données

Selon l'objectif défini et le public ciblé, un ensemble de données sera nécessaire pour soutenir le processus de compréhension (c'est-à-dire la création de sens) des utilisateurs visés. La portée des données doit être précisée, car elle est indépendante du public. Elle peut concerner un apprenant, la classe (ou une cohorte entière) ou même l'institution, mais elle

peut aussi concerner l'enseignant. L'inclusion de l'enseignant dans cette dimension favorise l'investigation de l'analyse des enseignants (Bienkowski *et al.*, 2012), où l'objectif est d'améliorer le processus d'apprentissage lui-même.

Tableau 2 : Objectifs exprimés liés au focus et au niveau de SA

Objectif	Focus	Niveau de SA
Améliorer la rétention et les résultats	Processus cognitif	Sensibilisation
Fournir une visualisation des performances d'apprentissage avec une comparaison de l'ensemble de la classe	Performance du processus d'apprentissage	Compréhension
Fournir des commentaires sur les activités et les performances d'apprentissage des étudiants	Performance du processus d'apprentissage	Action
Suivre les interactions des apprenants dans les systèmes d'apprentissage en ligne	Comportement d'apprentissage	Compréhension
Favoriser la réflexion et la prise de conscience de leur activité	Processus méta-cognitif	Sensibilisation
Permettre aux étudiants l'autorégulation de leurs apprentissages	Processus méta-cognitif	Compréhension
Visualiser l'évolution des relations entre les participants dans les forums de discussion, pour aider les étudiants à voir comment ils contribuent au groupe	Aspect social du processus d'apprentissage social	Sensibilisation
Identifier et prendre en charge les étudiants à risque	Gestion des personnes	Sensibilisation → Action

La source et la portée des données sont également des caractéristiques importantes de la conception. Les données peuvent être collectées en classe (analyse de classe), à partir de systèmes de gestion de l'apprentissage classiques, de l'institution (transcriptions du curriculum, profils des apprenants, ...) ou d'autres sources (analyse multimodale). La durée peut être aussi courte qu'une session ou aussi longue qu'une vie. La liste des différentes données nécessaires doit également être établie pour définir le tableau de bord. Dans cette dimension, les données peuvent provenir des environnements d'apprentissage eux-mêmes ou de transformations, telles que des agrégations ou des calculs d'indicateurs de haut niveau. Ces transformations sont cruciales pour convertir les données brutes en informations significatives et utiles pour les utilisateurs finaux.

2.4.6. Comment ? Visualisation et création de sens dans les TBA

La question du « Comment ? » est cruciale pour les tableaux de bord, car des visualisations adéquates sont nécessaires pour obtenir des aperçus significatifs (Chatti *et al.*, 2020 ; Sarikaya *et al.*, 2019). Comme identifié dans la section « Pourquoi », le tableau de bord peut soutenir un processus de création de sens, où l'utilisateur(s) peut naviguer à travers différents

niveaux de conscience de la situation et peut avoir besoin d'analyser des données. Les visualisations jouent un rôle clé dans la création de sens dans les tableaux de bord, en permettant aux utilisateurs de naviguer à travers différents niveaux de conscience et d'analyser des données. Elles peuvent prendre la forme de graphiques, de diagrammes ou de cartes et peuvent être interactives, permettant aux utilisateurs de manipuler les données et de découvrir des aperçus significatifs.

En matière d'interactions, Shneiderman (2003) propose une taxonomie concise pour guider les interactions avec des informations graphiques. La taxonomie de recherche d'informations, connue sous le nom de « Mantra de recherche d'information visuelle » (*Visual Information-Seeking Mantra*), englobe des tâches telles que fournir une vision globale des données, examiner une partie spécifique avec sélection et interaction, réduire la quantité de données, obtenir des détails précis, visualiser des relations, conserver un historique d'actions, et extraire une sous-collection ciblée.

2.4.7. Cadre de conception

L'ensemble des dimensions identifiées constitue un cadre exhaustif destiné à guider le processus de conception participative. Ce cadre se structure en deux parties distinctes (Figure 3). La première s'attache à détailler le contexte d'utilisation, englobant les trois premières questions. La seconde concerne la spécification des données et de leurs visualisations, en lien avec le focus, ou objet central du TBA. Par ailleurs, elle analyse les interactions nécessaires pour interagir avec le TBA afin de faciliter la création de sens. Les dimensions spécifiquement liées à la création de sens en vue de la prise de décision sont mises en évidence par l'ajout d'une étoile. En résumé, ce cadre offre une vision détaillée des éléments à spécifier dans un processus de conception des TBA, en intégrant de manière intrinsèque les aspects cruciaux liés à la création de sens.

3. OUTILLAGE DE LA CONCEPTION PARTICIPATIVE

Dans cette section, nous introduisons les outils que nous avons élaborés pour soutenir la méthodologie de conception participative et générative, en accord avec le cadre théorique DEFLAD. Notre approche repose sur une collaboration active avec les parties prenantes et s'appuie sur les dimensions de conception définies par DEFLAD. Pour instrumenter et renforcer cette démarche, nous avons conçu deux outils complémentaires qui facilitent l'exploration et la concrétisation des idées issues des différentes dimensions de conception : un outil d'idéation collaborative centré sur l'utilisateur et un outil de prototypage rapide destiné aux développeurs travaillant en collaboration avec les utilisateurs. En les intégrant dans le processus global de conception, notre objectif est de créer une synergie étroite entre les utilisateurs et les développeurs, facilitant ainsi la création de TBA mieux adaptés aux exigences du terrain, et intégrant le savoir-faire métier des développeurs. Ces outils sont spécifiquement conçus pour répondre aux étapes de *Create* et *Deliver* du modèle de co-design en analytique de l'apprentissage (Prieto-Alvarez *et al.*, 2018) (voir Figure 1).

3.1. OUTILLAGE DE LA PHASE D'IDÉATION

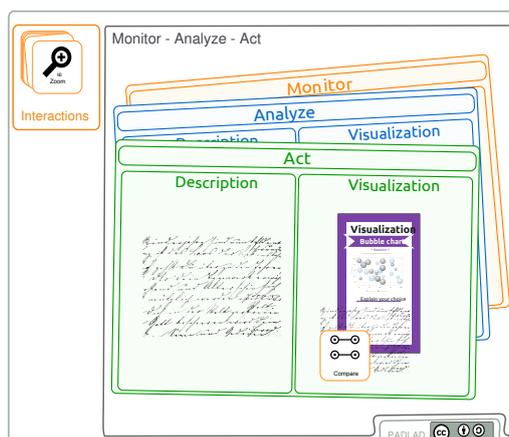
La phase d'idéation est une étape collaborative essentielle qui mobilise activement les parties prenantes, chacune contribuant avec ses perspectives uniques et ses connaissances spécialisées. Cette étape est cruciale dans le processus de conception, car elle permet de définir les objectifs de la conception de manière précise et exhaustive dès le début du processus. Pour soutenir efficacement cette phase, nous avons développé PaDLAD (Sadallah *et*

al., 2022), une boîte à outils d'idéation spécialement conçue pour cela. Cet ensemble d'outils, qui comprend des personas, des cartes d'exploration, des panneaux et des supports de sketching, vise à stimuler la créativité et à clarifier les différentes dimensions de l'espace de conception.



(a) Panneau d'identification

(b) Panneau DataViz



(c) Panneau de sketching

Figure 4 : Différents panneaux composant la boîte à outils PaDLAD

Destinée aux différents acteurs éducatifs, PaDLAD permet la création de tableaux de bord qui suivent non seulement les indicateurs associés aux objectifs, mais détectent également des situations potentiellement indésirables, identifient les preuves de ces situations, et rassemblent les informations nécessaires pour planifier des actions correctives. L'activité de conception avec cet outil est définie comme un processus en trois phases essentielles : (1) *Identification* de l'objectif du TBA ; (2) Définition des *données* et des *visualisations* appropriées ; et (3) *Sketching* en réalisant des croquis du TBA visé. Ces phases sont concrétisées par trois panneaux dédiés (Figure 4). Chaque panneau est conçu pour accueillir des artefacts spécifiques (cartes et formulaires) fournissant des informations liées à l'étape spécifique du processus de conception.

PANNEAU D'IDENTIFICATION. Ce panneau (Figure 4a) est dédié à la description du contexte d'utilisation spécifique du tableau de bord. Il comporte trois sections réservées pour accueillir une *formulaire de persona*, une *fiche d'objectif*, et un ensemble de *cartes de contexte*.

Le *formulaire de persona* est conçu pour personnifier les parties prenantes et collecter des informations pertinentes, telles que l'expertise, la culture visuelle, ou la motivation. Il offre ainsi une clarification sur l'utilisateur et le contexte du TBA.

Étant donné que l'objectif joue un rôle central dans le processus d'idéation, une *fiche d'objectif* est réservée pour soutenir son expression. Cette fiche permet d'énoncer le problème que le tableau de bord cherche à résoudre et la manière dont cette résolution doit être atteinte. L'objectif est défini en fonction du niveau souhaité de conscience de la situation et peut être adapté aux profils des participants, exprimant leurs besoins de manière différenciée, qu'ils soient apprenants, enseignants, ou autres.

Quatre *cartes de contexte* supplémentaires sont proposées pour fournir des informations relatives à l'utilisation anticipée du tableau de bord. Les *Cartes de Public* aident à définir le champ d'analyse, notamment les utilisateurs et la circulation de l'information entre eux. Les *Cartes de Données* précisent la source de données visée ainsi que le moment de la collecte des données (temps d'observation).

PANNEAU DATAVIZ (DONNÉES & VISUALISATIONS). Ce panneau (Figure 4b), axé sur la question *Quoi ?*, offre un espace spécifique pour les tuples composés de cartes de données et de visualisation. Les cartes *Données* (ou *Mesures*) identifient les données et indicateurs pertinents pour atteindre les objectifs du TBA. Les cartes *Visualisations* forment un ensemble de cartes technologiques proposant des représentations visuelles classiques.

L'interaction entre les cartes *Données* et *Visualisations* est cruciale. Les tuples sont construits de manière à définir les données et à les associer aux visualisations appropriées, assurant ainsi une représentation significative des informations contenues dans les cartes de données. Par exemple, un tuple pourrait associer une *Carte de Données* décrivant les taux de réussite des apprenants avec une *Carte de Visualisations* sous forme de diagramme à barres pour représenter ces taux. Un autre tuple pourrait lier les *Données* concernant l'engagement des utilisateurs avec la *Visualisation* d'un graphique en secteurs, offrant ainsi une vue rapide des activités prédominantes. Ainsi, ce panneau vise à créer une relation cohérente entre les données sélectionnées et les moyens visuels utilisés pour les présenter, favorisant une visualisation claire et pertinente dans le TBA.

LE PANNEAU SKETCHING. Ce panneau (Figure 4c) vise principalement à faciliter la création de vues et de représentations graphiques au cours du processus de conception. Les participants utilisent des esquisses pour donner forme au TBA, en combinant divers tuples de données/visualisations et en définissant des options d'interaction. Cette étape, à la fois pratique et visuelle, enrichit significativement le processus de conception.

Nous envisageons un TBA comme une interface interactive dynamique qui évolue en fonction des interactions de l'utilisateur. Dans cette optique, les vues générées ne sont pas statiques, mais sont plutôt façonnées et modifiées par les actions des utilisateurs. Trois catégories de vues sont distinguées, correspondant aux niveaux de conscience de situation que les utilisateurs cherchent à associer au TBA : (1) les vues *Perception*, qui sont orientées vers la surveillance de l'environnement et fournissent une vue globale des données pertinentes ; (2) les vues *Compréhension*, qui intègrent des représentations visuelles pour éclairer l'analyse et la compréhension de situations spécifique ; et (3) les vues *Projection*, qui préparent l'utilisateur à agir sur les situations découvertes et analysées dans les niveaux précédents. Ces vues sont matérialisées par des fiches contenant des tuples de données/visualisation placées sur le panneau. Il convient de souligner que toutes ces vues ne sont pas nécessairement présentes dans un TBA, et qu'un TBA peut inclure plusieurs vues du même type. Cette flexibilité assure une conception qui répond aux besoins spécifiques des utilisateurs. Ainsi, le TBA peut

être personnalisé et adapté pour fournir les informations les plus pertinentes et utiles pour chaque utilisateur ou situation.

L'interaction joue un rôle central dans la transition entre les différentes vues, permettant une progression graduelle de la *Perception* à l'analyse, puis à la *Compréhension*, et enfin à la *Projection*. Pour façonner ces interactions, nous nous sommes inspirés du *Mantra de la recherche d'information visuelle* de Shneiderman (2003), que nous avons intégré à l'espace de conception. Nous avons ainsi identifié six types d'interaction : (1) *Vue d'ensemble* permet à l'utilisateur d'avoir une vue globale des données disponibles ; (2) *Zoom* peut être exécuté pour examiner une partie des données en permettant de sélectionner cette partie ; (3) *Filtre-Recherche* réduit la quantité de données et/ou d'objets visuels affichés, et aide l'utilisateur à trouver et à se concentrer sur des éléments d'intérêt spécifiques ; (4) *Détails* peut être utilisé pour obtenir plus de précision sur les données ou une partie des données afin d'obtenir un meilleur aperçu ; (5) *Relation-Association-Comparaison* permet aux utilisateurs de visualiser les relations entre les points de données ; et (6) *Changement de vue* peut être utilisé pour changer le point d'intérêt de l'utilisateur.

Ce panneau de sketching guide les participants de trois manières distinctes dans la définition d'options visant à soutenir la création de sens par l'utilisateur du TBA. En premier lieu, les fiches associées aux différents niveaux de conscience de la situation sont différenciées par des couleurs spécifiques (*rouge* pour la surveillance, *bleu* pour l'analyse et la compréhension, et *vert* pour la projection et l'action). Ensuite, le panneau intègre des cartes technologiques pour aider les participants à exprimer toutes les options d'interaction avec le TBA et ses différentes vues. Enfin, chaque fiche est couplée à un formulaire encourageant les participants à décrire leur vision du processus de création de sens. Cette approche complète offre une diversité d'outils visuels et conceptuels, facilitant une exploration approfondie du potentiel du TBA tout en permettant la mise en place d'options pour renforcer la création de sens par l'utilisateur.

3.2. OUTILLAGE DE LA PHASE DE PROTOTYPAGE ET DE SUPPORT

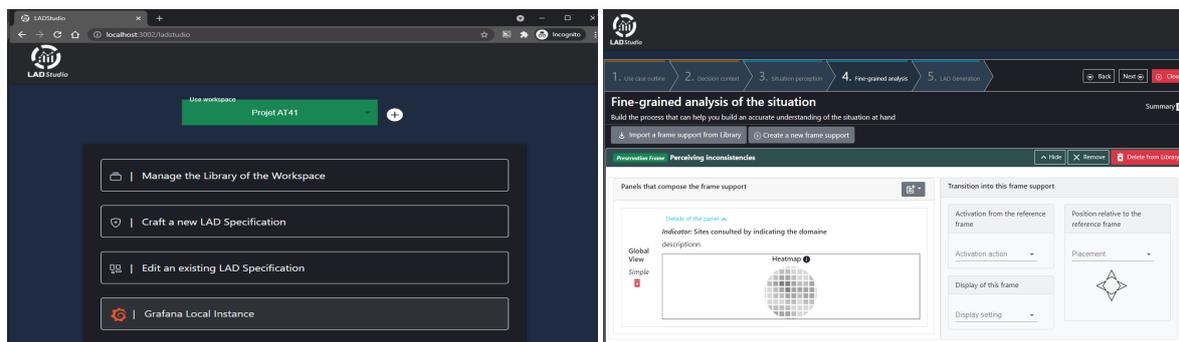
Pour soutenir la phase de prototypage, nous avons créé LADStudio (Sadallah et Gilliot, 2023), un outil qui permet de générer des prototypes de TBA via une spécification déclarative résultant de la phase d'idéation. La conception générative implique que les utilisateurs décrivent leurs besoins et intentions, explorent l'espace de conception, génèrent des solutions cibles, puis sélectionnent et affinent la solution la plus appropriée. Cette approche guide la conception progressive du TBA, impliquant les utilisateurs et les concepteurs à différentes étapes. Les utilisateurs participent à la description des besoins, à l'évaluation du résultat, tandis que les concepteurs dirigent la génération, explorent des options et évaluent les résultats avec les utilisateurs.

LADStudio est conçu selon une architecture à trois niveaux dans laquelle la présentation, le traitement des applications et la gestion sont des processus logiquement séparés. Nous l'avons implémenté en utilisant des technologies modernes. Le rendu du tableau de bord est testé avec une instance de Grafana¹, un outil open source d'analyse et de visualisation interactive. L'interface principal de LADStudio ainsi que ses composants essentiels sont illustrés sur la Figure 5 et seront examinés en détail dans les paragraphes suivants.

MODÈLE À BASE DE COMPOSANTS. Nous avons adopté une approche basée sur les composants pour modéliser la structure des tableaux de bord. De cette manière, la structure (et la vue logique) d'un tableau de bord est élaborée en utilisant une hiérarchie de composants

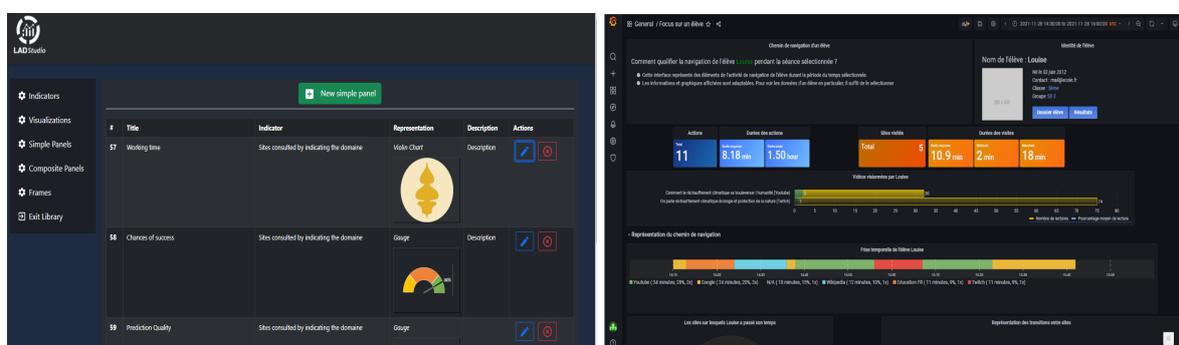
1. <https://grafana.com/grafana/>

imbriqués. Cette méthode facilite la conception des TBA en offrant des composants communs, partageables et réutilisables. De nouveaux composants peuvent être créés en adaptant des composants existants ou en utilisant des composants de niveau inférieur.



(a) Accueil et modules LADStudio

(b) Spécification d'un TBA



(c) Librairie de panneaux

(d) Prototype d'un TBA généré

Figure 5 : Principaux composants de LADStudio

Les composants d'un TBA se déclinent en plusieurs catégories (Figure 6) :

- *Indicateurs* : ces éléments de données représentent des mesures d'intérêt stratégique. Chaque indicateur est défini par un nom, une description qui en précise la signification et l'importance, ainsi qu'une formule de calcul qui en détaille la méthode de détermination.
- *Visualisations* : ces composants graphiques transforment les données en une représentation visuelle intuitive, facilitant ainsi l'analyse et l'interprétation par l'utilisateur. Elles peuvent prendre diverses formes, telles que des graphiques, des tableaux ou des cartes, et sont conçues pour mettre en évidence des tendances ou des modèles spécifiques dans les données.
- *Panels* : éléments structurants du tableau de bord, ils servent de conteneurs pour les visualisations et établissent des relations structurantes entre les composants. On distingue deux types de panels :
 1. les *panels de base*, qui associent les indicateurs aux visualisations correspondantes, permettant ainsi une interprétation directe et simplifiée des mesures clés ;
 2. les *panels composites*, qui combinent plusieurs panels de base pour créer des structures plus complexes, offrant une vue d'ensemble et une analyse comparative des différents indicateurs.
- *Vues* : ce sont des composants de haut niveau conçus pour soutenir les différents niveaux de conscience de la situation. Ils fournissent les éléments structurels et dé-

finissent l’interactivité nécessaire pour développer une compréhension approfondie de l’environnement. Trois types de vues sont identifiés, chacun correspondant à un niveau différent de conscience de la situation :

1. les vues *Perception*, qui permettent à l’utilisateur de surveiller son environnement à travers des indicateurs et des visualisations immédiatement perceptibles ;
2. les vues *Compréhension*, qui offrent des représentations aidant l’utilisateur à analyser et à comprendre une situation spécifique grâce à des visualisations interactives et des panels de base ;
3. les vues *Projection*, qui préparent l’utilisateur à intervenir sur les situations identifiées et analysées dans les niveaux précédents, en utilisant des panels composites pour anticiper les tendances futures et planifier des actions.

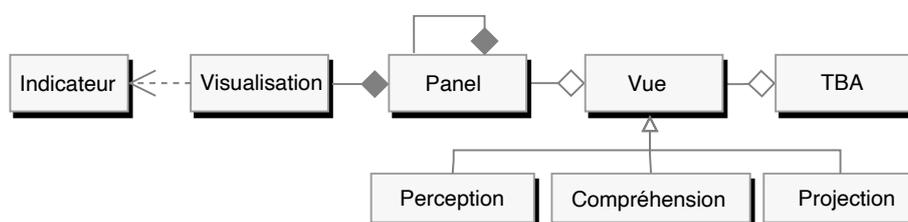


Figure 6 : Structure à composants des tableaux de bord

MODÈLE D’INTERACTION. Afin de définir les types d’interaction qui favorisent la création de sens, nous avons intégré six catégories d’interaction, explicitement détaillées dans DEFLAD, décrites dans PaDLAD, et inspirées du *Mantra de la recherche d’information visuelle* de Shneiderman (2003). L’interaction de l’utilisateur avec le composant cible (panneau ou vue) déclenche des processus en arrière-plan qui mettent à jour le TBA. Cela peut entraîner l’affichage de nouvelles données que l’utilisateur peut intégrer à ses connaissances internes en construisant un cadre mental (*framing*) ou en modifiant son cadre actuel, voire en consolidant ce dernier.

BIBLIOTHÈQUE DES COMPOSANTS DU TBA. Dans l’objectif de favoriser la réutilisation et la collaboration, la *Bibliothèque des Composants* offre une plateforme permettant de définir, de stocker et de partager divers éléments cruciaux pour la composition de tableaux de bord (voir Figure 5c). Cette bibliothèque regroupe cinq catégories de composants inhérents au modèle TBA (voir Figure 6) : (1) les *Modèles d’indicateurs*, définissant les critères de suivi ; (2) les *Visualisations*, proposant des représentations graphiques des données ; (3) les *Panels Simples*, constitués d’éléments autonomes ; (4) les *Panels Composites*, regroupant plusieurs panels simples pour une vue consolidée ; et (5) les *Vues*, déterminant les différentes perspectives du TBA.

L’utilisateur a la possibilité d’enrichir cette bibliothèque en concevant de nouveaux composants à partir de rien ou en utilisant des éléments préexistants. En outre, il a la capacité de modifier les composants existants ou, si nécessaire (par exemple, pour éviter la redondance), de les supprimer. Chaque composant est intrinsèquement lié à des interactions spécifiques, ce qui soutient le processus global de création de sens. Cette méthode favorise non seulement la flexibilité dans la conception des TBA, mais aussi la formation d’une communauté collaborative qui partage des ressources essentielles pour une conception optimale.

ASSISTANT DE SPÉCIFICATION DE TBA. Ce composant de l'outil permet de spécifier un tableau de bord étape par étape (Figure 5b). Le scénario de spécification se compose de cinq écrans séquentiels : (1) description du contexte cible et de l'utilisation visée ; (2) définition de l'objectif principal du tableau de bord ; (3) description des vues de suivi permettant de surveiller l'environnement, par rapport à l'objectif défini ; (4) description des vues d'analyse et de compréhension permettant d'atteindre les niveaux de conscience de situation : compréhension et projection, en lien avec l'aspect surveillé ; et (5) génération d'une spécification du tableau de bord.

Le prototype généré peut être visualisé sur une instance Grafana embarquée ou sur une installation distante (voir l'exemple présenté sur la Figure 5d). Une spécification de tableau de bord produite avec LADStudio peut être rééditée, définissant un processus cyclique d'édition et de test. De plus, les composants (indicateurs, visualisations, panneaux et vues) produits sont automatiquement sauvegardés, ce qui simplifie l'alimentation de la bibliothèque de LADStudio pour la réutilisation des composants.

4. EXPÉRIMENTATIONS

4.1. PROBLÉMATIQUE

Dans cette section, nous rapportons les expérimentations que nous avons réalisées pour évaluer les outils proposés comme support à la conception participative de TBA. Nos questions de recherche sont les suivantes :

- 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 ? (*QR1*)
- L'outil de conception participative encourage-t-il la collaboration et soutient-il la créativité participative au sein d'un groupe ? (*QR2*)
- L'outil de prototypage assure-t-il une bonne expérience utilisateur ? (*QR3*)

4.2. MÉTHODOLOGIE

Nous avons mené deux études qualitatives distinctes pour évaluer nos outils. Pour l'évaluation de l'outil d'idéation PaDLAD, nous avons sollicité la participation de quinze individus, comprenant des enseignants, des chercheurs et des concepteurs pédagogiques sans expertise spécifique en EIAH. D'autre part, pour l'évaluation de l'outil de prototypage LADStudio, nous avons impliqué un autre groupe de treize participants. Ces derniers étaient également des enseignants, des chercheurs et des concepteurs pédagogiques, mais possédaient des connaissances techniques sur les tableaux de bord et le développement d'interfaces liées aux EIAH. Il est important de noter que, bien que LADStudio ait été initialement destiné aux développeurs, des participants avec des compétences techniques ont été recrutés en raison de la difficulté à trouver un nombre suffisant de développeurs de TBA pour participer à l'étude. Les données socio-démographiques des participants sont présentées dans le Tableau 3.

4.3. PROCÉDURE

Pour évaluer PaDLAD, un atelier de conception participative de TBA a été organisé. L'objectif était de construire collectivement des TBA adaptés au contexte du collège, répondant aux besoins des enseignants et des ingénieurs pédagogiques. Les chercheurs ont introduit l'outil lors d'une phase d'introduction, exposant les objectifs de l'atelier et détaillant chaque aspect de la boîte à outils. Ils ont ensuite guidé les discussions en veillant à ce que les instructions et le protocole soient respectés, et en répondant aux questions des participants.

Après avoir exposé le protocole de mise en œuvre de la conception participative, les participants ont été répartis en groupes équilibrés pour favoriser une diversité de perspectives. Chaque groupe a disposé d'environ une heure et demie pour travailler sur sa conception.

Tableau 3 : 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

Pour l'évaluation de LADStudio, la procédure a été similaire. Après une session de présentation et de démonstration de l'outil, les participants ont été invités à expérimenter l'outil individuellement pendant une courte période, puis à participer à un atelier de conception participative similaire à celui de PaDLAD. Le chercheur a continué de jouer un rôle de facilitateur dans chaque étape, guidant les discussions, expliquant les outils et les méthodes, veillant à ce que les instructions soient suivies, et répondant aux questions des participants. L'étude a duré environ deux heures.

À la fin de chaque étude, les participants ont été invités à remplir individuellement un questionnaire évaluant divers aspects de l'outil, y compris l'utilisabilité, la qualité de la participation et l'expérience utilisateur.

4.4. INSTRUMENTS

Nous avons suivi une approche qualitative, utilisant des questionnaires et recueillant les réactions des participants. Ces questionnaires sont fondés sur des instruments validés que nous avons modifiés pour les adapter au contexte français. Pour garantir leur pertinence et leur compréhension, nous avons soumis ces questionnaires à l'arbitrage de collègues chercheurs et de deux concepteurs pédagogiques expérimentés. Nous avons ensuite révisé les questionnaires en fonction des recommandations reçues.

UTILISABILITÉ. La facilité d'utilisation de PaDLAD et celle de LADStudio ont été évaluées avec les participants à la fin de chaque étude en remplissant des questionnaires SUS (System Usability Scale) (Brooke *et al.*, 1996) dédiés. Nous avons décidé d'utiliser le SUS, qui est un test d'utilisabilité standardisé, pour trois raisons principales (1996 ; Tullis et Stetson, 2004) : (1) il couvre une variété d'aspects de la convivialité, tels que le besoin d'assistance, la facilité d'apprentissage et la complexité ; (2) il peut être utilisé sur un petit échantillon tout en fournissant des résultats fiables ; et (3) il s'est avéré être un instrument valide pour évaluer efficacement la convivialité des systèmes. Sur la base du modèle général du SUS (Brooke *et al.*, 1996) et de sa version française (F-SUS) (Gronier et Baudet, 2021), nous avons conçu deux questionnaires, contenant chacun les dix énoncés SUS. Chaque affirmation impaire était formulée sur un ton positif, tandis que les affirmations paires étaient formulées sur un ton négatif. Le niveau d'accord était mesuré sur une échelle de Likert, allant de *Tout à fait en désaccord* (noté 1) à *Tout à fait en accord* (noté 5).

QUALITÉ DE LA PARTICIPATION. Pour évaluer le degré d'implication des participants lors des ateliers de conception, nous avons utilisé la *Self-Report Level of Participation Survey* (SRLPS) proposé par Hyett *et al.* (2020), instrument adapté d'un outil de mesure de la participation, le *spidergram* (Draper *et al.*, 2010). Le SRLPS se concentre sur cinq activités de codesign, que nous définissons dans notre contexte comme suit :

- *Planification et organisation* : cette activité implique la détermination des aspects du TBA à aborder à chaque étape de la conception. Cela peut inclure la définition des objectifs, l'établissement d'un calendrier et la répartition des tâches.
- *Processus de conception créative* : il s'agit d'un ensemble d'activités qui favorisent l'émergence d'idées et de constructions de conception innovantes pour le TBA. Cela peut inclure le brainstorming, le dessin de croquis, la modélisation, etc.
- *Définition des priorités* : cette activité consiste à identifier les priorités pour les tâches de conception du TBA et à déterminer les éléments de la conception à mettre en avant ou à atténuer. Cela peut impliquer des discussions de groupe, des votes, etc.
- *Négociation* : il s'agit d'une discussion visant à explorer différentes options de conception du TBA dans le but de parvenir à un accord. Cela peut impliquer des compromis, des discussions ouvertes, etc.
- *Réflexion et évaluation* : cette activité consiste en une session de brainstorming avec le groupe pour évaluer la pertinence de chaque choix de conception du TBA par rapport à l'objectif fixé. Cela peut inclure des discussions de groupe, des retours d'information, etc.

Nous avons invité les participants à évaluer individuellement leur niveau de participation pour chaque activité. Les notes attribuées étaient les suivantes : 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).

EXPÉRIENCE UTILISATEURS. L'évaluation de l'expérience utilisateur (UX) donne un aperçu du niveau de confort d'une personne par rapport à la satisfaction d'un système, et détermine les domaines à améliorer. Nous avons utilisé le questionnaire sur l'expérience de l'utilisateur UEQ (User Experience Questionnaire), un outil valide qui sert à évaluer de manière exhaustive l'expérience utilisateur des produits interactifs (Laugwitz *et al.*, 2008), applicable à de petits groupes (Schrepp *et al.*, 2014). Le questionnaire regroupe un total de 26 items répartis en six échelles (Santoso *et al.*, 2016) :

- *Attraction* décrit l'impression générale que les utilisateurs avaient de l'outil ;
- *Efficacité* qualifie la possibilité d'utiliser l'outil rapidement et efficacement ;
- *Compréhensibilité* décrit la facilité avec laquelle il est possible de comprendre comment utiliser l'outil et de s'y familiariser ;
- *Fiabilité* qualifie le sentiment de l'utilisateur d'avoir le contrôle de son interaction et d'être en confiance avec l'outil ;
- *Stimulation* décrit si l'utilisation de l'outil est excitante et motivante ; *Originalité* décrit dans quelle mesure la conception de l'outil est innovante et créative, et attire l'attention de l'utilisateur.

4.5. RÉSULTATS

4.5.1. Évaluation de l'utilisabilité des outils

Afin d'analyser les résultats de l'étude d'utilisabilité, nous avons calculé les scores SUS en suivant la procédure proposée par Brooke *et al.* (1996). Ces scores sont calculés en utili-

sant l'équation présentée dans la Formule 1, où n est le nombre de participants, $m = 10$ est le nombre de questions, q_{ij} est le score individuel par question et par participant, et $norm = 2.5$ est un facteur de normalisation.

$$\overline{SUS} = \frac{1}{n} \sum_{i=1}^n norm. \sum_{j=1}^m \begin{cases} q_{ij} - 1, & \text{si } q_{ij} \bmod 2 > 0 \\ 5 - q_{ij}, & \text{sinon} \end{cases} \quad (1)$$

Le score SUS est une échelle qui s'étend de 0 à 100. Un score de 0 signifie que l'utilisateur considère le système comme totalement inutile, tandis qu'un score de 100 signifie que l'utilisateur considère le système comme extrêmement utile. Tout score dépassant 68 est considéré comme étant au-dessus de la moyenne. Dans Bangor *et al.* (2009), les scores SUS sont mis en correspondance avec une échelle d'adjectifs afin d'attacher une signification plus descriptive au score SUS attribué à un système. Selon Brooke *et al.* (1996), le SUS devrait être considéré comme une mesure unidimensionnelle de l'utilisabilité. Par conséquent, dans l'analyse de nos résultats, nous nous concentrerons sur le score global SUS plutôt que sur les scores individuels des questions.

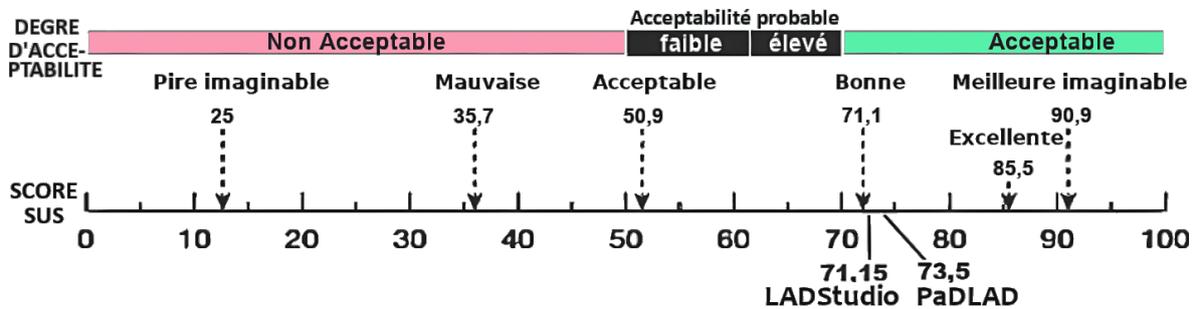


Figure 7 : Résultats de l'étude d'utilisabilité de PaDLAD et de LADStudio

Les résultats de l'évaluation des participants pour les deux outils sont présentés sur la Figure 7. La valeur totale du score SUS obtenu pour PaDLAD était de 1102,5, avec une valeur moyenne de 73,5, un écart-type de 9,34 et une médiane de 72,5. En utilisant le classement par grade, et comme le montre Figure 7, le score SUS de 73,5 attribué à la boîte à outils signifie que le niveau de l'utilisateur *Plage d'acceptabilité* est élevé, que le niveau de l'échelle des grades est *catégorie C* et que le niveau de l'évaluation des adjectifs de l'utilisateur est *Catégorie bonne*. Cela indique un niveau satisfaisant et acceptable d'utilisabilité de PaDLAD par les participants.

L'évaluation de LADStudio a donné un score SUS total de 925 avec une valeur moyenne de 71,15, un écart-type de 6,15 et une médiane de 72,5. En utilisant ce classement par grade proposé par Bangor *et al.* (2009) (voir Figure 7), le score SUS de l'outil de 71,15 signifie que le niveau de l'utilisateur *Plage d'acceptabilité* est *Acceptable*, le niveau de l'échelle d'adjectifs *catégorie C*, et le niveau de l'utilisateur *Catégorie adjective* est *Bon*. Cela suggère que les participants ont trouvé l'outil suffisamment utilisable et acceptable.

4.5.2. Qualité de participation de PaDLAD

Pour analyser les données recueillies à l'aide du questionnaire SRLPS, tel que préconisé par ses auteurs (Hyett *et al.*, 2020), nous avons converti les niveaux de participation en valeurs numériques. Dans cette échelle, 1 correspond à une participation passive (niveau bas), 2 au partage d'informations, 3 à l'engagement et la mobilisation, 4 à la collaboration, et 5 à l'autonomisation (niveau élevé). Le Tableau 4 présente la médiane et l'intervalle des scores pour chaque activité clé mesurée par l'enquête.

Tableau 4 : 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.

Bien qu'il y ait une certaine variabilité dans les scores entre les participants individuels, la note cible de collaboration (médiane de 4) a été atteinte pour toutes les activités. Les activités liées à la négociation des choix de conception et au processus de conception créative ont reçu le plus haut niveau de participation autodéclarée. Cela démontre l'efficacité de la boîte à outils pour créer un environnement propice à la communication, à l'échange et à la créativité lors des ateliers de conception participative. Cependant, malgré un score plutôt satisfaisant, l'activité associée à la définition des priorités a enregistré le niveau de participation moyen le plus bas. Cela pourrait indiquer un besoin d'orienter davantage les participants et de simplifier l'outil proposé et le protocole associé. Enfin, il est intéressant de noter que bien qu'aucun des participants n'ait qualifié sa participation de passive, aucun d'entre eux n'a non plus indiqué le plus haut niveau de participation pour aucune des activités. Cela pourrait s'expliquer par le fait que l'expérience de co-design est nouvelle pour eux, et qu'ils peuvent se sentir 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.

Ces résultats mettent en avant le potentiel de PaDLAD pour stimuler une participation élevée. Cependant, ils soulignent également la nécessité d'un accompagnement constant des participants durant tout le processus de conception, en particulier pour la définition des priorités. De plus, ils illustrent l'importance d'approfondir l'étude des facteurs susceptibles de limiter la participation des participants à son niveau le plus élevé.

4.5.3. Expérience utilisateur de LADStudio

Nous avons calculé les résultats de l'UEQ en utilisant les outils proposés par Laugwitz *et al.* (2008) après avoir échelonné les réponses des participants de -3 (extrême négatif) à +3 (extrême positif) sur une échelle de Likert. Les scores allant de -0,8 à 0,8 reflètent une évaluation neutre de la dimension correspondante, tandis que les scores supérieurs à 0,8 indiquent une évaluation positive et ceux inférieurs à 0,8 impliquent une évaluation négative.

Comme le montre la représentation graphique des résultats présentée sur la Figure 8, la note globale est suffisamment élevée. Le score moyen le plus élevé est celui de *Originalité*, avec une moyenne de 2,48 (écart-type = 0,17), suivi par *Efficacité* (moyenne = 2,27, écart-type = 0,90) et *Stimulation* (moyenne = 2,25, SD = 0,18). Ces scores sont d'un excellent niveau. La fiabilité a un bon score (moyenne = 1,64, écart-type = 0,17). Le résultat le moins positif concerne la dimension *Compréhensibilité* (moyenne = 1,13, écart-type = 0,73), ce qui signifie que les participants ont éprouvé quelques difficultés à comprendre l'utilisation de l'outil. Les scores *Originalité*, *Efficacité* et *Stimulation* sont à un excellent niveau, ce qui signifie que les participants ont considéré LADStudio comme très créatif, efficace et stimulant. Les résultats ont également montré que *Compréhensibilité* avait un résultat mitigé. Cela

peut s'expliquer par le temps relativement court que les participants passent sur l'outil par rapport à la quantité d'informations dont ils disposent. Nous sommes également conscients que l'adoption d'un tel outil nécessite un certain effort de diffusion et de formation auprès des concepteurs et de tous ses utilisateurs potentiels. LADStudio est un prototype et nécessite quelques itérations supplémentaires pour améliorer l'intuitivité de ses fonctionnalités.

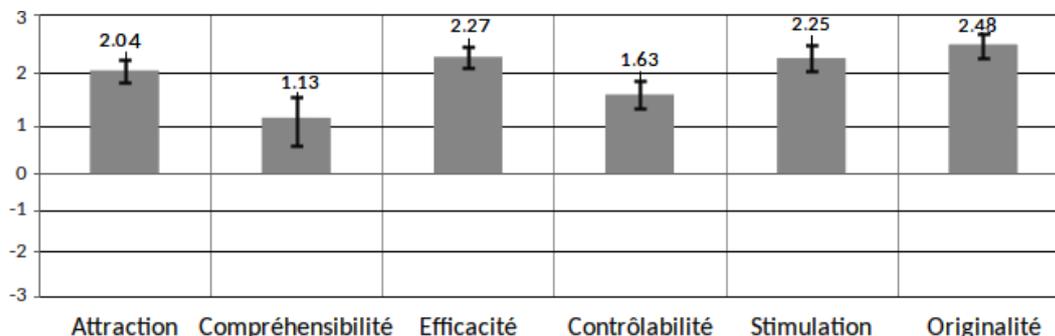


Figure 8 : Résultats de l'évaluation de l'expérience utilisateur de LADStudio

4.5.4. Commentaires des participants

Les commentaires des participants ont été recueillis à deux moments clés du processus : d'abord après les ateliers de conception participative avec PaDLAD, puis après des sessions de prototypage avec LADStudio. Ces séances ont permis aux participants de donner leurs impressions sur chaque outil et de partager leurs expériences tout au long du processus de conception. Les commentaires ont été consignés par écrit lors de séances de débriefing qui ont suivi ces activités, permettant ainsi une documentation précise des retours des participants et une compréhension approfondie de leurs expériences.

RETOURS ET COMMENTAIRES SUR PADLAD. Les participants ont apprécié la démarche globale de conception participative de PaDLAD, soulignant son rôle stimulant dans la création d'un débat passionné et la génération d'idées originales pour améliorer l'apprentissage en classe. Un participant a exprimé : « *L'approche participative de PaDLAD a vraiment encouragé un échange d'idées dynamique, ce qui a été crucial pour explorer des solutions innovantes aux défis éducatifs* ». L'outil PaDLAD, utilisé lors de la phase d'idéation, a été apprécié pour sa capacité à susciter des idées de conception originales, favorisées par la dynamique d'échange et de partage. Un autre participant a partagé : « *Les sessions d'idéation avec PaDLAD ont été particulièrement enrichissantes. L'emploi de cet outil a considérablement facilité la collaboration, nous permettant ainsi de générer des idées créatives de manière efficace* ».

Un défi majeur rencontré pendant l'idéation a été le processus de négociation nécessaire pour établir un persona, mettant en lumière les divers traits de personnalité, défis et aspirations des participants. Certains ont soulevé des interrogations sur le niveau de détail nécessaire dans la description du persona. Un participant a partagé son point de vue : « *La création du persona a été un défi, surtout pour déterminer le niveau de détail. Plus d'exemples pratiques auraient été utiles* ».

Une fois le persona établi, les groupes ont défini des objectifs, parfois déviés de ceux initialement préconçus, soulignant ainsi la nécessité d'une flexibilité dans la conception des tableaux de bord. Un participant a noté : « *La flexibilité dans la définition des objectifs a été*

bénéfique. Cela nous a permis d'explorer des idées différentes et de trouver des solutions adaptées à nos besoins spécifiques ».

Bien que la spécification claire du tableau d'identification ait facilité la construction de l'image cible, les débats sur les représentations visuelles appropriées ont souligné le besoin de soutien à ce stade, notamment en raison des différents niveaux de littératie visuelle. Plusieurs participants ont exprimé leur besoin d'orientations plus claires sur le choix des représentations visuelles en fonction des données. Un participant a mentionné : *« Le choix des visualisations peut être délicat. Des conseils sur la correspondance entre les données et les représentations visuelles seraient utiles ».*

En résumé, il apparaît que renforcer la littératie en données et en visualisation est crucial pour une adoption plus efficace de PaDLAD. Néanmoins, cet outil s'est avéré utile pour surmonter la difficulté de concevoir des TBA malgré une culture limitée dans ces domaines. La construction des vues en fonction des niveaux de conscience a été particulièrement appréciée pour sa capacité à refléter les étapes du raisonnement et à permettre une projection dans des scénarios d'utilisation réels. Un participant a souligné : *« 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 ».*

Cependant, un autre participant a soulevé des défis liés à la construction des panneaux associés aux niveaux de conscience de la situation : *« La possibilité de projeter le processus de raisonnement dans des représentations visuelles 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 ».* Certains participants ont partagé leurs expériences personnelles sur la manière dont ils ont réussi à simplifier ce processus, tandis que d'autres ont évoqué la nécessité d'outils complémentaires pour rendre cette étape plus accessible. Un participant a suggéré : *« Intégrer des modèles préconçus pourrait rendre la construction des panneaux plus simple pour ceux qui débutent ».*

RETOURS ET COMMENTAIRES SUR LADSTUDIO. Les participants ont apprécié la stratégie de LADStudio pour relever les défis associés à la conception de TBA. Certains ont exprimé leur satisfaction quant à la capacité de l'outil à clarifier les attentes et à concrétiser les réponses. Un participant a souligné : *« Un tel outil permet de ne pas rester au niveau des idées et des croquis, mais de passer à des prototypes fonctionnels concrets, ce qui est essentiel pour concrétiser les concepts ».*

L'utilisation de LADStudio a été perçue comme nécessitant une compréhension approfondie du contexte et de l'objectif de l'outil. Un participant a partagé son expérience : *« J'ai réussi à utiliser l'outil correctement après un certain temps de pratique, mais il est essentiel de comprendre comment il s'inscrit dans le processus global de conception. C'est une sorte de transition des idées à des prototypes fonctionnels ».* Les participants ont également souligné l'accent positif mis par LADStudio sur la conception de TBA soutenant le processus de prise de sens de l'utilisateur. Un participant a partagé son point de vue : *« J'apprécie l'effort déployé pour concevoir des TBA qui soutiennent vraiment la manière dont les utilisateurs créent du sens. Cela donne une perspective très intéressante et innovante ».*

Plusieurs participants ont mentionné la facilité d'utilisation du module de spécification progressive de LADStudio pour décrire un TBA. Un participant a noté : *« La possibilité de décrire progressivement un TBA à travers des écrans séquentiels simplifie le processus de spécification, permettant ainsi une meilleure itération et test ».* En ce qui concerne la bibliothèque de composants, les participants ont apprécié son rôle dans la simplification de l'alimentation et la possibilité de capitaliser et réutiliser des composants. Un participant a partagé : *« La bibliothèque de composants est un ajout utile, facilitant la réutilisation et la*

capitalisation des composants créés. Cela rend le processus de conception plus efficace ».

Cependant, certains participants ont signalé des difficultés similaires à celles de PaDLAD, notamment concernant l'utilisation de concepts théoriques liés au processus décisionnel. Selon un des participant : « *L'intégration de ces concepts peut être un peu complexe au début, ce qui pourrait être un obstacle pour certains utilisateurs lors de leurs premières interactions avec l'outil. Il serait bénéfique d'avoir des guides plus spécifiques sur ces concepts* ».

En résumé, bien que LADStudio ait été perçu comme un outil novateur pour résoudre des défis spécifiques, les commentaires des participants soulignent la nécessité d'une meilleure compréhension du contexte, d'une période d'apprentissage initiale, d'une attention particulière à l'intégration de concepts théoriques, ainsi que d'une appréciation positive des fonctionnalités telles que le module de spécification progressive et la bibliothèque de composants. Ces éléments sont cruciaux pour améliorer l'expérience globale des utilisateurs et favoriser une utilisation plus efficace de l'outil dans des contextes de conception de TBA.

4.6. DISCUSSION

Les évaluations révèlent que les participants ont utilisé les outils sans rencontrer de difficultés significatives, ce qui témoigne d'une perception positive de leur attrait et de leur facilité d'utilisation. Ceci 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 *QR3*.

L'évaluation de l'utilisabilité, mesurée par le questionnaire SUS, atteste d'une bonne performance de la boîte à outils PaDLAD, avec un score global de 73,5, situé dans une fourchette d'acceptabilité élevée. De manière similaire, LADStudio affiche une bonne utilisabilité avec un score SUS moyen de 71,15. Les résultats de l'UEQ confirment quant à eux une expérience utilisateur globalement positive, établissant ainsi la qualité des aspects évalués. Parallèlement, l'analyse du niveau de participation des intervenants souligne l'efficacité de la boîte à outils et de sa méthodologie associée pour encourager une collaboration fructueuse. Ces constats sont étayés par les retours positifs des participants, ce qui est de bon augure pour le développement futur de PaDLAD et de LADStudio.

Néanmoins, la discussion se doit d'explorer les limites inhérentes à ces outils. La méthode par carte, bien que novatrice, expose des défis pratiques, avec des utilisateurs éprouvant parfois des difficultés à manipuler certaines composantes de la boîte à outils. Ces entraves ont eu des répercussions sur la compréhension des informations synthétisées sur les différentes cartes. Pour pallier ces obstacles, il est impératif de fournir des instructions détaillées, des directives précises, des exemples illustratifs, et d'instaurer un encadrement attentif avec un soutien personnalisé. Cette nécessité souscrit à l'idée que la documentation exhaustive de la boîte à outils et le rôle du facilitateur doivent être intégrés à la conception (Grudin, 1994). De plus, la méthodologie qualitative employée expose des biais bien identifiés, tels que le biais d'auto-sélection, la pensée de groupe et l'influence des personnalités dominantes (Silverman et Marvasti, 2008). Le biais d'auto-sélection, particulièrement saillant dans notre étude, compromet la généralisation des résultats à une population plus vaste. L'admission franche de ces limitations souligne l'importance d'adopter une approche critique dans l'interprétation des résultats (Bryman, 2016). Pour renforcer la validité externe, l'expansion de la taille de l'échantillon est préconisée (Nielsen, 2000). Bien que des études antérieures confirment la fiabilité du SUS et de l'UEQ avec des groupes restreints (Schrepp *et al.*, 2014 ; Tullis et Stetson, 2004), élargir la base d'utilisateurs renforcerait la robustesse des conclusions.

Enfin, il est impératif d'étendre la portée de l'évaluation au-delà de la convivialité. L'intégration d'autres aspects de l'expérience utilisateur, au-delà de la facilité d'utilisation, revêt une importance cruciale. La réalisation d'une enquête qualitative complémentaire, conjointement avec l'étude sur le niveau de participation (SRLPS), permettrait d'obtenir une compréhension approfondie de la variabilité des scores individuels. Une étude longitudinale de plus grande envergure permettrait une évaluation plus détaillée de la qualité des tableaux de bord construits, offrant ainsi une perspective sur l'impact de ces outils sur la conception pédagogique des TBA. Cette approche est en accord avec la recommandation d'une évaluation à long terme pour mesurer l'efficacité des technologies éducatives. Comme le soulignent Parry *et al.* (2018), il est important de prendre en compte la complexité, l'évolution dans le temps et l'interaction avec le contexte local lors de l'évaluation des initiatives d'amélioration. Ces principes peuvent être directement appliqués à l'évaluation des technologies éducatives.

5. CONCLUSION

L'impératif de placer l'utilisateur au centre du processus de conception se confirme comme une exigence cruciale pour le développement de produits véritablement adaptés à des besoins spécifiques (Vilpola, 2008). Notre contribution vise à renforcer cette approche dans le domaine complexe de la conception des TBA, en mettant un accent particulier sur l'amélioration de leur adoption, souvent entravée. Dans cette optique, nous avons élaboré le cadre DEFLAD, une proposition spécifique qui guide la création des TBA en intégrant des capacités de création de sens dans le processus décisionnel. En cohérence avec DEFLAD, notre espace de conception des TBA intègre des éléments liés à la dimension de la création de sens, facilitant une décomposition précise des objectifs visés avec des niveaux de conscience de la situation. Les résultats des études révèlent l'intérêt de nos propositions destinées à accroître l'adoption des outils d'analyse de l'apprentissage. L'approche participative, soutenue par des outils spécialement conçus, a reçu un accueil favorable de la part des parties prenantes, surmontant avec succès deux obstacles majeurs : le manque d'implication des utilisateurs finaux et la complexité technique inhérente à la création des TBA.

Notre méthodologie de conception engage activement les parties prenantes dans le processus, autorisant les concepteurs à générer rapidement des prototypes conformes aux besoins des utilisateurs finaux. Pour soutenir ce processus participatif, nous avons introduit PaDLAD, une boîte à outils spécifique à la conception des TBA construite autour du cadre DEFLAD et de son espace de conception. Cette boîte à outils, rassemblant divers instruments, facilite la créativité partagée et permet l'expression des besoins des utilisateurs au travers de profils de personas, de cartes d'idéation et de croquis. L'utilisation de cartes structurées accélère le raffinement des idées, offrant une représentation persistante des échanges entre les participants et capturant la dimension cruciale de la création de sens.

Pour la phase de prototypage et de test, nous avons déployé LADStudio, aligné sur l'espace de conception DEFLAD, afin de fournir un appui concret à la prise de conscience et de décision des utilisateurs finaux. LADStudio offre une plateforme où les utilisateurs peuvent concrétiser les concepts générés pendant la phase d'idéation. Cette phase cruciale s'appuie sur la simplification de l'approche et des outils pour favoriser un développement génératif de prototypes fonctionnels. En mettant l'accent sur la convivialité et l'accessibilité, LADStudio permet aux utilisateurs de créer des prototypes interactifs de TBA de manière intuitive.

En conclusion, la collecte de propositions de TBA auprès des parties prenantes, facilitée par l'utilisation combinée de PaDLAD et de LADStudio, permet non seulement de recueillir des besoins supplémentaires, mais également d'identifier de nouveaux usages potentiels pour les TBA. Grâce à la méthodologie participative, les utilisateurs ont la possibilité d'expri-

mer leurs besoins spécifiques et de contribuer activement à la conception, ce qui enrichit la compréhension des scénarios d'utilisation et ouvre des perspectives pour des applications innovantes des TBA dans divers contextes éducatifs.

Les études menées ont confirmé l'innovation de l'approche et des outils développés, démontrant leur capacité à surmonter avec succès les obstacles majeurs à l'adaptation des TBA. Ces résultats ouvrent de nouvelles perspectives de recherche, notamment en ce qui concerne la transférabilité des TBA entre différents contextes d'utilisation et les invariants entre propositions issues de groupes d'utilisateurs variés, visant à répondre aux besoins diversifiés de la communauté d'apprentissage. Nous sommes convaincus que la capitalisation des connaissances et l'exploration de l'assistance à la conception participative représentent des axes de développement prometteurs pour l'avenir des TBA. En tirant parti des éléments produits lors du processus de conception et en facilitant la collaboration entre les utilisateurs, nous pouvons optimiser l'efficacité et la pertinence des TBA. De plus, en explorant davantage les possibilités d'assistance à la conception participative, nous pouvons rendre les outils encore plus accessibles et conviviaux, favorisant ainsi leur adoption et leur utilisation dans une variété de contextes éducatifs. Enfin, nous lançons un appel à une collaboration plus étroite avec la communauté pour approfondir la compréhension des dimensions de conception et leur application pratique dans le contexte des outils ayant pour objectif de soutenir l'éducation. En travaillant ensemble, chercheurs et praticiens peuvent continuer à innover et à faire progresser ces outils pour répondre aux besoins évolutifs des apprenants et des enseignants.

RÉFÉRENCES

- Alhadad, S. S. (2018). Visualizing data to support judgement, inference, and decision making in learning analytics : Insights from cognitive psychology and visualization science. *Journal of Learning Analytics*, 5(2), 60-85.
- Alhamadi, M. (2020). Challenges, strategies and adaptations on nteractive dashboards. *Proceedings of the 28th ACM Conference on user modeling, adaptation and personalization*, 368-371.
- Alvarez, C. P., Martinez-Maldonado, R., et Shum, S. B. (2020). LA-DECK : A card-based learning analytics co-design tool. *Proceedings of the tenth international conference on learning analytics & knowledge*, 63-72.
- Bangor, A., Kortum, P., et Miller, J. (2009). Determining what individual SUS scores mean : Adding an adjective rating scale. *Journal of usability studies*, 4(3), 114-123.
- Bienkowski, M., Feng, M., et Means, B. (2012). *Enhancing teaching and learning through educational data mining and learning analytics : An issue brief* (rapp. tech.). Office of Educational Technology, US Department of Education.
- Bodily, R., et Verbert, K. (2017). Review of research on student-facing learning analytics dashboards and educational recommender systems. *IEEE Transactions on Learning Technologies*, 10(4), 405-418.
- Boscardin, C., Fergus, K. B., Hellevig, B., et Hauer, K. E. (2018). Twelve tips to promote successful development of a learner performance dashboard within a medical education program. *Medical Teacher*, 40(8), 855-861.
- Brooke, J., et al. (1996). SUS-A quick and dirty usability scale. *Usability evaluation in industry*, 189(194), 4-7.
- Bryman, A. (2016). *Social research methods*. Oxford University Press.
- Campos, F. C., Ahn, J., DiGiacomo, D. K., Nguyen, H., et Hays, M. (2021). Making sense of sensemaking : Understanding how k-12 teachers and coaches react to visual analytics. *Journal of Learning Analytics*, 1-21.

- Chatti, M. A., Muslim, A., Guliani, M., et Guesmi, M. (2020). The LAVA model : Learning analytics meets visual analytics. Dans *Adoption of data analytics in higher education learning and teaching* (p. 71-93). Springer.
- DiSalvo, B., Yip, J., Bonsignore, E., et DiSalvo, C. (2017). *Participatory design for learning : Perspectives from practice and research*. Taylor & Francis.
- Dollinger, M., Liu, D., Arthars, N., et Lodge, J. M. (2019). Working together in learning analytics towards the co-creation of value. *Journal of Learning Analytics*, 6(2), 10-26.
- Draper, A. K., Hewitt, G., et Rifkin, S. (2010). Chasing the dragon : developing indicators for the assessment of community participation in health programmes. *Social science & medicine*, 71(6), 1102-1109.
- Echeverria, V., Martinez-Maldonado, R., Granda, R., Chiluiza, K., Conati, C., et Shum, S. B. (2018). Driving data storytelling from learning design. *Proceedings of the 8th International Conference on Learning Analytics & Knowledge*, 131-140.
- Endert, A., Hossain, M. S., Ramakrishnan, N., North, C., Fiaux, P., et Andrews, C. (2014). The human is the loop : new directions for visual analytics. *Journal of Intelligent Information Systems*, 43(3), 411-435.
- Endsley, M. (1995). Toward a theory of situation awareness in dynamic systems : Situation awareness. *Human Factors*, 37(1), 32-64.
- Ez-Zaouia, M., Tabard, A., et Lavoué, E. (2020). Emodash : A dashboard supporting retrospective awareness of emotions in online learning. *International Journal of Human-Computer Studies*, 139, 102411.
- Few, S. (2006). *Information dashboard design : The effective visual communication of data*. O'Reilly Media, Inc.
- Garrison, D. R., Anderson, T., et Archer, W. (2003). A theory of critical inquiry in online distance education. *Handbook of Distance Education*, 1(4), 113-127.
- Gašević, D., Dawson, S., et Siemens, G. (2015). Let's not forget : Learning analytics are about learning. *TechTrends*, 59(1), 64-71.
- Gaver, B., Dunne, T., et Pacenti, E. (1999). Design : cultural probes. *Interactions*, 6(1), 21-29.
- Gilliot, J.-M., Iksal, S., Medou, D. M., et Dabbebi, I. (2018). Conception participative de tableaux de bord d'apprentissage. *IHM'18 : 30e Conférence Francophone sur l'Interaction Homme-Machine*, pp-119.
- Gilliot, J.-M., et Sadallah, M. (2024). A framework for co-designing effective LADs supporting sensemaking and decision making. *International Journal of Learning Technology*, 19(1), 109-130.
- Gronier, G., et Baudet, A. (2021). 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.
- Grudin, J. (1994). *Computer-supported cooperative work : History and focus*. Academic Press.
- Holstein, K., McLaren, B. M., et Aleven, V. (2017). Intelligent tutors as teachers' aides : Exploring teacher needs for real-time analytics in blended classrooms. *Proceedings of the 7th International Learning Analytics and Knowledge Conference*, 257-266.
- Hutchins, E. (1995). *Cognition in the Wild*. MIT Press.
- Hyett, N., Bagley, K., Iacono, T., McKinstry, C., Spong, J., et Landry, O. (2020). 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.

- Ifenthaler, D., et Widanapathirana, C. (2014). Development and validation of a learning analytics framework : Two case studies using support vector machines. *Technology, Knowledge and Learning*, 19(1), 221-240.
- Jivet, I., Scheffel, M., Drachler, H., et Specht, M. (2017). Awareness is not enough : Pitfalls of learning analytics dashboards in the educational practice. Dans É. Lavoué, H. Drachler, K. Verbert, J. Broisin et M. Pérez-Sanagustín (dir.), *Data Driven Approaches in Digital Education* (p. 82-96). Springer, Cham.
- Klein, G., Moon, B., et Hoffman, R. R. (2006). Making sense of sensemaking 1 : Alternative perspectives. *IEEE Intelligent Systems*, 21(4), 70-73.
- Laugwitz, B., Held, T., et Schrepp, M. (2008). Construction and evaluation of a user experience questionnaire. *Symposium of the Austrian HCI and usability engineering group*, 63-76.
- Liu, Z., Nersessian, N., et Stasko, J. (2008). Distributed cognition as a theoretical framework for information visualization. *IEEE Transactions on Visualization and Computer Graphics*, 14(6), 1173-1180.
- Lucero, A., Dalsgaard, P., Halskov, K., et Buur, J. (2016). Designing with Cards. Dans *Collaboration in Creative Design* (p. 75-95). Springer.
- Meyer, J., Thomas, J., Diehl, S., Fisher, B., et Keim, D. A. (2010). From visualization to visually enabled reasoning. *Dagstuhl Follow-Ups*, 1.
- Ndukwe, I. G., et Daniel, B. K. (2020). Teaching analytics, value and tools for teacher data literacy : A systematic and tripartite approach. *International Journal of Educational Technology in Higher Education*, 17(1), 1-31.
- Newell, A., et Simon, H. A. (1972). *Human problem solving* (T. 104). Prentice-Hall Englewood Cliffs, NJ.
- Nielsen, J. (2000). Why you only need to test with 5 users.
- Park, Y., et Jo, I.-H. (2015). Development of the learning analytics dashboard to support students' learning performance. *Journal of Universal Computer Science*, 21(1), 110.
- Parry, G., Coly, A., Goldmann, D., Rowe, A. K., Chattu, V., Logiudice, D., Rabrenovic, M., et Nambiar, B. (2018). Practical recommendations for the evaluation of improvement initiatives. *International Journal for Quality in Health Care*, 30(suppl_1), 29-36.
- Pirolli, P., et Card, S. (2005). The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. *Proceedings of International Conference on Intelligence Analysis*, 5, 2-4.
- Prieto-Alvarez, C. G., Martinez-Maldonado, R., et Anderson, T. D. (2018). Co-designing learning analytics tools with learners. Dans J. Lodge, J. Horvath et L. Corrin (dir.), *Learning Analytics in the Classroom* (p. 93-110). Routledge.
- Robertson, T., et Simonsen, J. (2012). Participatory Design : an introduction. Dans *Routledge international handbook of participatory design* (p. 1-17). Routledge.
- Russell, D. M., Stefik, M. J., Pirolli, P., et Card, S. K. (1993). The cost structure of sensemaking. *Proceedings of the INTERACT'93 and CHI'93 Conference on Human Factors in Computing Systems*, 269-276.
- Sadallah, M., et Gilliot, J.-M. (2023). Generating LADs that make sense. *Proceedings of the 15th International Conference on Computer Supported Education - Volume 1 : CSEDU 2023*, 35-46.
- Sadallah, M., Gilliot, J.-M., Iksal, S., Quelennec, K., Vermeulen, M., Neyssensas, L., Aubert, O., et Venant, R. (2022). Designing LADs that promote sensemaking : A participatory tool. Dans I. Hilliger, P. J. Muñoz-Merino, T. De Laet, A. Ortega-Arranz et T. Farrell (dir.), *Educating for a New Future : Making Sense of Technology-Enhanced Learning Adoption* (p. 587-593). Springer International Publishing.

- Sanders, E. B.-N., et Stappers, P. J. (2008). Co-creation and the new landscapes of design. *CoDesign*, 4(1), 5-18.
- Santoso, H. B., Schrepp, M., Isal, R., Utomo, A. Y., et Priyogi, B. (2016). Measuring user experience of the student-centered e-learning environment. *Journal of Educators Online*, 13(1), 58-79.
- Sarikaya, A., Correll, M., Bartram, L., Tory, M., et Fisher, D. (2019). What do we talk about when we talk about dashboards? *IEEE Transactions on Visualization and Computer Graphics*, 25(1), 682-692.
- Sarmiento, J. P., et Wise, A. F. (2022). Participatory and co-design of learning analytics : An initial review of the literature. *Proceedings of the 12th International Learning Analytics and Knowledge Conference*, 535-541.
- Schrepp, M., Hinderks, A., et Thomaschewski, J. (2014). Applying the user experience questionnaire (UEQ) in different evaluation scenarios. *International Conference of Design, User Experience, and Usability*, 383-392.
- Schuler, D., et Namioka, A. (1993). *Participatory design : Principles and practices*. CRC Press.
- Schulz, H.-J., Hadlak, S., et Schumann, H. (2011). The design space of implicit hierarchy visualization : A survey. *IEEE Transactions on Visualization and Computer Graphics*, 17(4), 393-411.
- Schwendimann, B. A., Rodríguez-Triana, M. J., Vozniuk, A., Prieto, L. P., Boroujeni, M. S., Holzer, A., Gillet, D., et Dillenbourg, P. (2017). Perceiving learning at a glance : A systematic literature review of learning dashboard research. *IEEE Transactions on Learning Technologies*, 10(1), 30-41.
- Sedrakyan, G., Mannens, E., et Verbert, K. (2019). Guiding the choice of learning dashboard visualizations : Linking dashboard design and data visualization concepts. *Journal of Computer Languages*, 50, 19-38.
- Shaw, M. (2012). The role of design spaces. *IEEE software*, 29(1), 46-50.
- Shneiderman, B. (2003). The eyes have it : A task by data type taxonomy for information visualizations, 364-371.
- Silverman, D., et Marvasti, A. (2008). *Doing qualitative research : A comprehensive guide*. Sage.
- Taffe, S. (2015). The hybrid designer/end-user : Revealing paradoxes in co-design. *Design Studies*, 40, 39-59.
- Tullis, T. S., et Stetson, J. N. (2004). A comparison of questionnaires for assessing website usability. *Usability Professionals Association (UPA) 2004 Conference*.
- Verbert, K., Ochoa, X., De Croon, R., Dourado, R. A., et De Laet, T. (2020). Learning analytics dashboards : the past, the present and the future. Dans *Proceedings of the 10th International Conference on Learning Analytics & Knowledge* (p. 35-40).
- Vilpolo, I. H. (2008). A method for improving ERP implementation success by the principles and process of user-centred design. *Enterprise Information Systems*, 2(1), 47-76.
- Visser, F. S., Stappers, P. J., van der Lugt, R., et Sanders, E. B.-N. (2005). Contextmapping : experiences from practice. *CoDesign*, 1(2), 119-149.
- Wang, F., et Hannafin, M. J. (2005). Design-based research and technology-enhanced learning environments. *Educational Technology Research and Development*, 53(4), 5-23.
- Wise, A. F., et Jung, Y. (2019). Teaching with analytics : Towards a situated model of instructional decision-making. *Journal of Learning Analytics*, 6(2), 53-69.
- Khakaj, F., Aleven, V., et McLaren, B. M. (2016). How teachers use data to help students learn : Contextual inquiry for the design of a dashboard. *European Conference on Technology Enhanced Learning EC-TEL 2016*, 340-354.

Yoo, Y., Lee, H., Jo, I.-H., et Park, Y. (2015). Educational dashboards for smart learning : Review of case studies. Dans G. Chen, V. Kumar, S. Kinshuk, R. Huang et S. C. Kong (dir.), *Emerging Issues in Smart Learning. Lecture Notes in Educational Technology* (p. 145-155). Springer Berlin Heidelberg.

Vers la conception de moyens et méthodes fondés sur les modèles pour caractériser et diagnostiquer la maturité numérique des enseignants

Towards the design of model-based means and methods to characterize and diagnose teachers' digital maturity

Christine MICHEL¹ ; Laëtitia PIERROT²

¹Université de Poitiers, UR TECHNE, Poitiers, France

²Le Mans Université, UR CREN, Le Mans, France

Résumé. Cet article traite de la manière de combiner les modèles de maturité numérique des enseignants pour en proposer une version unifiée, utilisable pour concevoir des moyens et méthodes de diagnostic. Onze modèles applicables au champ de l'enseignement scolaire obligatoire ont été identifiés par une revue de littérature. Les modèles et la manière dont leurs dimensions constitutives contribuent à déterminer des niveaux de maturité ont été analysés. La synthèse réalisée met en lumière la diversité des dimensions utilisées, et une prise en compte partielle de la maturité numérique. De plus, la plupart de ces modèles se concentrent sur les derniers niveaux de maturité, associés à des enseignants innovateurs ou pionniers. Les modèles ont tendance à écarter les enseignants non-utilisateurs ou faibles utilisateurs numériques, pourtant représentés dans le contexte français. Dans la dernière partie de l'article, une proposition de modèle unifié de la maturité numérique des enseignants, MUME, répondant à ces deux problèmes, est décrite ainsi que les résultats préliminaires d'une étude visant à concevoir une méthode de diagnostic.

Mots-clés : modèle de maturité numérique, intégration du numérique à l'école, pratiques numériques des enseignants

Abstract. *This article examines how models of teacher digital maturity can be combined to produce a unified version that can be used to design diagnostic tools and methods. 11 models applicable to the field of compulsory education were identified through a literature review. The models and how their constituent dimensions contribute to the determination of maturity levels were analyzed. The summary highlights the diversity of the dimensions used and the fact that digital maturity is only partially taken into account. What's more, most of these models focus on the most recent maturity levels associated with innovative or pioneering teachers. The models tend to exclude teachers who are not digital users or who have a low level of digital use, but who are present in the French context. In the final part of the article, a proposal for a unified model of teachers' digital maturity, MUME, which addresses these two issues, is described, together with the preliminary results of a study aimed at designing a diagnostic method.*

Keywords: *technology maturity model, technology integration in education, teachers' practices*

1. INTRODUCTION

Dans un contexte d'émergence et de diffusion d'innovations technologiques, la transformation numérique est l'un des enjeux les plus critiques en éducation (Antonietti *et al.*, 2023). Ce processus exige des enseignants des capacités à s'adapter et être préparé à intégrer de nouveaux outils, constamment renouvelés (McCarthy *et al.*, 2023). En France, la transformation est d'autant plus cruciale que les rapports sur les usages des enseignants du premier comme du second degré témoignent d'une faible intégration des outils numériques¹. Le potentiel d'intégration des outils numériques dans 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 ces outils numériques sont utilisés pour stimuler cognitivement les élèves et les engager dans des activités d'apprentissage (Antonietti *et al.*, 2023). Et, bien que la crise sanitaire de 2020 ait eu un effet de stimulation sur les usages numériques, ils restent encore limités à des pratiques de communication et transmission de ressources (Michel et Pierrot, 2022 ; Plantard et Serreau, 2024).

Avec l'objectif de décrire le déploiement des technologies en éducation, différents modèles sont proposés. Par exemple, le TPACK (Mishra et Koehler, 2006), le SAMR (Puentedura, 2012), le NETS-T (ISTE, 2017) ou le DigCompEdu (Redecker, 2017) abordent les différentes dimensions de l'activité professionnelle d'un enseignant. Ces modèles traitent plus globalement des dynamiques d'intégration et des niveaux de maturité numérique des enseignants. Les niveaux de maturité numérique représentent un moyen de mesurer la manière dont la technologie est utilisée pour transformer et redéfinir les activités professionnelles de l'enseignant, au-delà de considérations quantitatives (fréquences d'usage par exemple). Ces niveaux tiennent en effet compte de la qualité de l'enseignement, comme le type d'activités mises en œuvre (Backfisch *et al.*, 2021).

Les modèles existants sont relativement hétérogènes dans leur manière de considérer la nature des activités professionnelles ou les niveaux de maturité. Par exemple, le modèle TPACK (Mishra et Koehler, 2006) décrit les différents champs de connaissances (pédagogique, didactique et technique) mobilisées par les enseignants, lorsqu'ils utilisent des technologies dans des activités de formation. Si ce modèle présente l'avantage majeur de proposer une conceptualisation sommaire de l'activité de formation, cette simplification ne permet pas de faire état de ses spécificités (par exemple, s'agit-il d'une activité d'apprentissage ? d'enseignement ? à quelles fins ? selon quel format ?), ni des habiletés précises attendues pour la mettre en œuvre. Cependant, les modèles identifiés sont comparables dans le sens où ils n'incluent pas de descriptions liées à la maîtrise de technologies particulières, de manière à pouvoir rester valides alors que l'innovation technologique est en constante évolution.

Dans cette perspective, un modèle de maturité numérique peut être mobilisé pour soutenir les individus dans le processus de transformation numérique (McCarthy *et al.*, 2023). En effet, un tel modèle décrit les différentes dimensions de l'activité professionnelle nécessaires ou touchées par le processus de transformation. La maturité numérique y est définie comme la capacité d'un individu à s'emparer des technologies de manière à soutenir son propre développement personnel et à s'intégrer socialement (Laaber *et al.*, 2023). Rapportée au contexte éducatif, la vision globale des modèles de maturité numérique peut guider les enseignants ou l'équipe d'encadrement dans le choix des pratiques professionnelles à

¹ <https://www.education.gouv.fr/media/95365/download>

développer. L'enjeu principal est alors de faire en sorte d'aligner ces pratiques relatives à l'utilisation de technologies à des objectifs de réussite scolaire (à travers l'implication des élèves ou la personnalisation des apprentissages par exemple). Dit autrement, les modèles de maturité numérique en éducation 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.

Au-delà des problématiques d'accès, de disponibilité et de fréquence d'usage, la maturité numérique doit prendre en considération, dans ces modèles, 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 (Franklin et Bolick, 2007). Les modèles de maturité trouvent aussi leur intérêt pour mesurer, diagnostiquer ou accompagner les enseignants dans leur utilisation du numérique (Kimmons *et al.*, 2020). Plus largement, l'étude de la maturité à travers des modèles permet d'aborder l'adoption des outils numériques en combinant des facteurs liés à l'enseignant, et à son contexte d'exercice professionnel (Harrison *et al.*, 2014). Pour cela, il est nécessaire de considérer l'apprenant, l'enseignant et le contexte d'utilisation plus large en collectant des données pour mesurer l'étendue et la profondeur de l'intégration du numérique dans un établissement (Underwood *et al.*, 2010 ; Underwood et Dillon, 2004). 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 (ISTE, 2017 ; Leite et Lagstedt, 2021). L'état de l'art réalisé dans le cadre de cet article s'inscrit dans cette perspective.

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+Edu².

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). Quels moyens et quelles méthodes mobiliser pour mesurer la maturité numérique des enseignants ? (QR3).

2. MÉTHODE

Nous avons travaillé selon une méthode de revue herméneutique (Sackstein *et al.*, 2022), qui consiste à développer une compréhension à travers une analyse itérative et une

² <https://pix.fr/actualites/actualite-pix-edu/>

interprétation de travaux existants sur un sujet particulier. Ce processus de revue comprend notamment la recherche, la classification, et l'évaluation critique de ces travaux (Boell et Cecez-Kecmanovic, 2014). Pour ce faire, nous avons identifié : les modèles de maturité des enseignants (1) en partant des précédentes revues systématiques de littérature portant sur les modèles d'intégration et de maturité numérique en éducation (Carvalho *et al.*, 2018 ; Franklin et Bolick, 2007 ; Harrison *et al.*, 2014 ; Kimmons *et al.*, 2020 ; Leite et Lagstedt, 2021 ; Solar *et al.*, 2013) et dans d'autres organisations (Pee et Kankanhalli, 2009 ; Teichert, 2019), (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 nouveau. Seuls les modèles applicables au contexte de l'enseignement obligatoire ont été retenus, soit 21 modèles, listés dans la figure 1 (Michel et Pierrot, 2023).



Figure 1 : Liste des 21 modèles identifiés

Nous avons ensuite comparé ces modèles en considérant plusieurs critères (Kimmons *et al.*, 2020 ; Teichert, 2019) : *le périmètre* associé à l'application du modèle (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, présentés dans la suite de l'article, 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.

3. 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. Cinq 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.). 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.

Tableau 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 (BECTA, 2008)	G	G	F	O	Di	E	N
CIT (Leit et Lagstedt, 2021)	G	G	F	N	A	E	N
DigCompEdu (Redecker, 2017)	S	G	P	O	A	T	O
ICAP (Chi et al., 2018)	G	P	P	O	A	T	O
ICTE-MM (Solar et al., 2013)	G	G	P	O	A	E	O
LoTI (Moersch, 1995 ; Stoltzfus, 2006)	G	P	F	O	Di	T	O
NETS-T (ISTE, 2017)	G	G	P	N	A	E	O
PICRAT (Kimmons et al., 2020)	S	P	P	N	De	T	N
SAMR (Puentedura, 2012)	G	P	F	O	De	E	N
TIM (Kozdras et Welsh, 2018)	G	P	P	O	Di	E	O
TPACK (Mishra et Koehler, 2006)	G	P	F	F	D	T	O

4. LES MODÈLES DE MATURITÉ

Les modèles de maturité s'inscrivent dans des perspectives différentes, selon qu'ils décrivent les dynamiques d'appropriation (4.1), les dimensions mobilisées pour favoriser la maturité numérique (4.2), l'approche pédagogique mobilisée par les enseignants (4.3), la recherche d'efficacité pédagogique (4.4) et les compétences attendues (4.5) des enseignants ou qu'ils s'intéressent spécifiquement aux organisations (4.6).

4.1 MODÈLES BASÉS SUR LES DYNAMIQUES D'APPROPRIATION DES ENSEIGNANTS

Selon Puentedura (2012), le modèle Substitution Augmentation, Modification et Redéfinition -SAMR- encourage les éducateurs à « passer » des niveaux d'enseignement grâce à la technologie (voir figure 2a). Ce modèle fournit la structure pour considérer la manière dont l'intégration de l'outil numérique est réalisée à travers 4 étapes de transformation de la tâche d'enseignement : 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 (Blundell *et al.*, 2022).

Le modèle *Collective Integration of Technology* -CIT- (Leite et Lagstedt, 2021) considère le processus collectif de renforcement des connaissances d'un groupe (enseignants, responsables pédagogiques, décideurs) et la manière dont la culture de l'organisation peut soutenir (ou entraver) l'intégration des technologies éducatives dans les pratiques scolaires, dans une perspective interactionnelle (voir figure 2b). Le modèle identifie 4 états, plutôt qu'étapes, pour signaler que ces états, déclinés selon des aspects comportementaux, cognitifs et émotionnels ainsi que des expériences collectives, 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 dans des phases d'*autonomisation* (durant lesquelles ils construisent les premiers usages) et *exploration* (durant lesquelles ils développent de nouveaux usages).

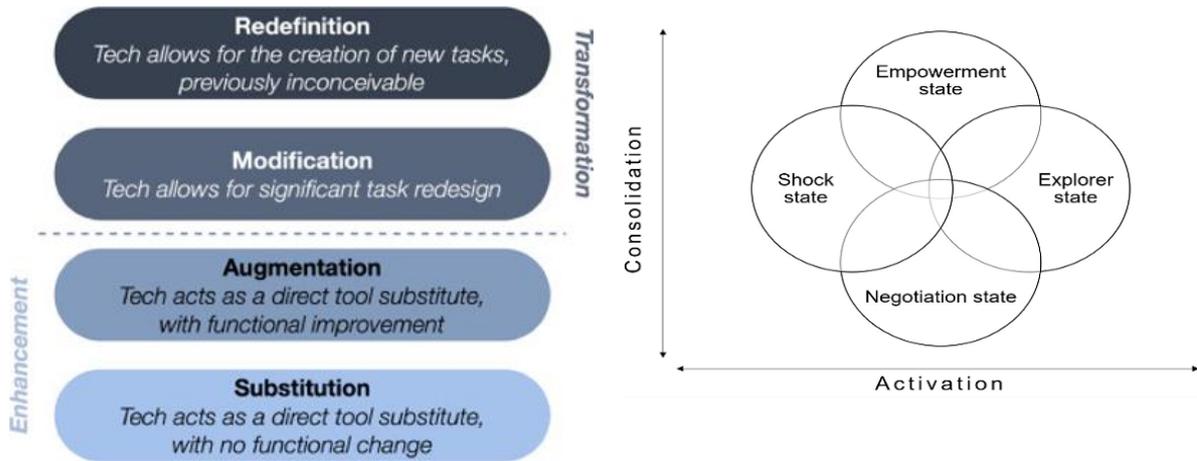


Figure 2 : (a) Modèle SAMR (Puentedura, 2012) et (b) CIT (Leite et Lagstedt, 2021)

4.2 MODÈLE BASÉ SUR LES DIMENSIONS DE LA MATURITÉ POUR LES ENSEIGNANTS

Le modèle *Technological Pedagogical Content Knowledge* -TPACK- (Mishra et Koehler, 2006) est considéré comme l'un des modèles les plus importants décrivant les compétences des enseignants pour un enseignement réussi avec la technologie (voir figure 3). La valeur ajoutée du modèle est de ne pas considérer individuellement les connaissances technologiques (*TK*), relatives au contenu disciplinaire (*CK*) et celles pédagogiques (*PK*), mais plutôt leurs interactions matérialisées par les zones de recouvrement (*TCK*, *PCK*, *TPK* et *TPACK*). En 2019, TPACK évolue pour inclure les connaissances contextuelles (*XK*) et ainsi intégrer les contraintes organisationnelles et situationnelles liées à la mise en œuvre de la technologie en enseignement (Mishra, 2019). 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.

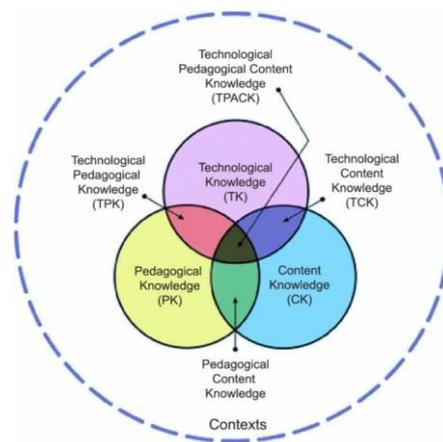


Figure 3 : Modèle TPACK avec la dimension XK (Mishra, 2019)

4.3 MODÈLES BASÉS SUR L'APPROCHE PÉDAGOGIQUE RETENUE PAR LES ENSEIGNANTS

Le modèle *Interactive, Constructive Active and Passive* -ICAP- (Chi *et al.*, 2018) propose une perspective différente en se concentrant sur les processus et niveaux d'engagement cognitifs des apprenants plutôt que sur le niveau de maturité ou la capacité des enseignants. Ce modèle (voir figure 4a) identifie quatre types d'activités d'apprentissage : *interactives*, *constructives*, *actives* et *passives*, en classant l'engagement cognitif du plus coûteux au moins coûteux (Antonietti *et al.*, 2023).

Les *Levels of Technology Implementation* (LoTi) (voir figure 4b) ont 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, le modèle LoTi décrit 5 dimensions (*enseignement/apprentissage avec le numérique, évaluation avec le numérique, créativité des élèves, développement professionnel et citoyenneté numérique*) dans une approche holistique, pour évaluer la portée complète de l'intégration des technologies. L'utilisation des outils et ressources numériques, en classe pour l'enseignement et l'apprentissage, est mesurée à l'aide d'outils validés empiriquement (Moersch, 1995 ; Stoltzfus, 2006) et vise à contribuer ensuite au développement professionnel des enseignants.

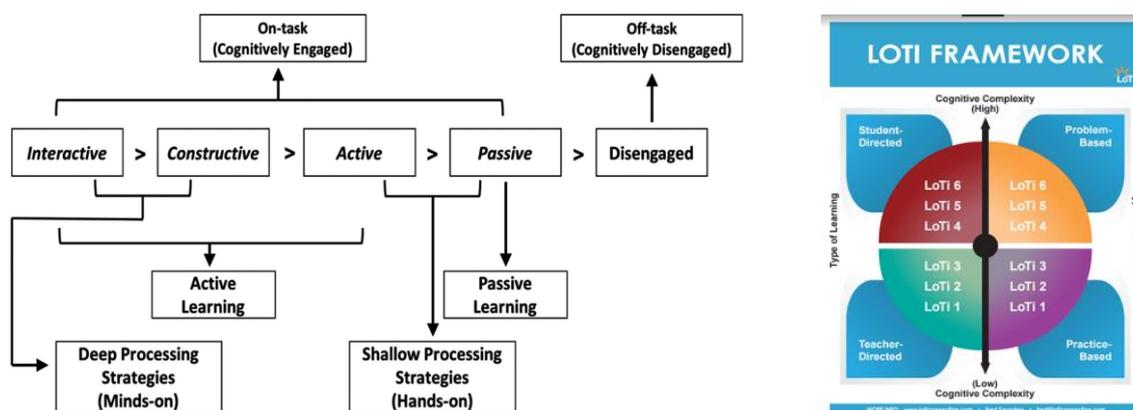


Figure 4 : (a) Modèle ICAP (Chi *et al.*, 2018) et (b) LoTi (Moersch, 1995 ; Stoltzfus, 2006)

4.4 MODÈLES MIXTES ARTICULANT L'EFFICACITÉ PÉDAGOGIQUE ET LES NIVEAUX DE MATURITÉ

Le modèle *Passive, Interactive, Creative, Replacement, Amplification, Transformation* (PICRAT) (Kimmons *et al.*, 2020) propose une approche centrée sur l'engagement de l'apprenant avec les outils numériques. Le modèle (voir figure 5a) se divise en deux composantes principales : l'engagement de l'apprenant avec l'outil (PIC pour *passif, interactif, créatif*) et la manière dont l'outil modifie la mise en œuvre pédagogique de l'activité (RAT pour *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 poursuivis.

décrits selon 4 indicateurs (ISTE, 2017). 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 (Crompton et Sykora, 2021).

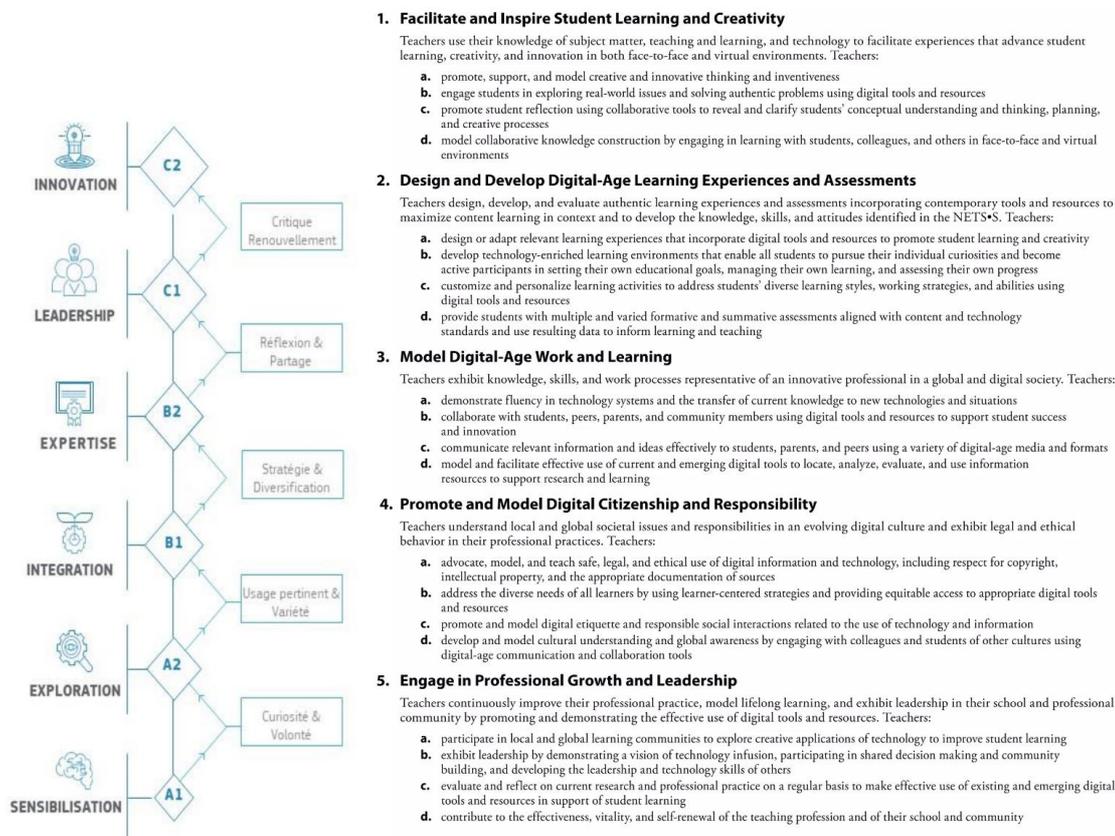


Figure 6 : (a) Modèle DigCompEdu (Redecker, 2017) et (b) NETS-T (ISTE, 2017)

4.6 MODÈLES DESCRIPTIFS DE LA MATURITÉ DES ORGANISATIONS

Le modèle BECTA (voir figure 7a) développé par le *British Educational Communications and Technology Agency* en 2008 est conçu pour aider les établissements d'enseignement supérieur à atteindre une maturité numérique à travers un outil d'auto-évaluation 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 (BECTA, 2008). Le modèle a été complété en 2018 (Ristić, 2018) 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.

ICT in School Education Maturity Model (ICTE-MM) est une proposition (voir figure 7b) qui a pour ambition de se rapprocher de standards internationaux (Solar *et al.*, 2013) et s'inspire pour cela du modèle standardisé du CMMI⁴ (*Capability Maturity Model Integration*) et des travaux de l'ISTE (2017). Le modèle distingue 3 dimensions susceptibles de soutenir les processus éducatifs (pilotage, stimulation de la culture numérique, ressources

⁴ <https://www.cmmiinstitute.com/>

d'information et TIC). ICTE-MM propose un outil d'auto-évaluation et une feuille de route pour guider les chefs d'établissement sur la gestion du numérique.

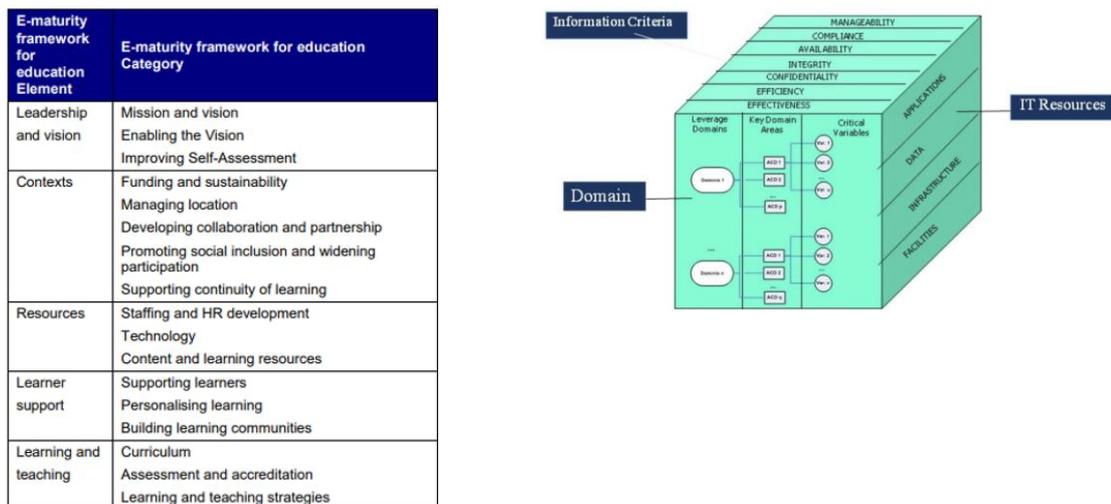


Figure 7 : (a) Modèle BECTA (BECTA, 2008) et (b) ICTE-MM (Solar *et al.*, 2013)

5. VERS UN MODÈLE UNIFIÉ DE LA MATURITÉ NUMÉRIQUE DES ENSEIGNANTS : MUME

5.1 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 chaque modèle. Différentes restructurations ont été opérées de manière à articuler les modèles entre eux dans une vue unifiée (voir figure 8b). 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 (voir figure 8a) : le TPACK, le ICTE-MM, le DigCompEdu et l'ICAP. Nous avons choisi de ne faire apparaître dans le TPACK que les dimensions qui concernent l'intégration des technologies, soit TPACK 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.

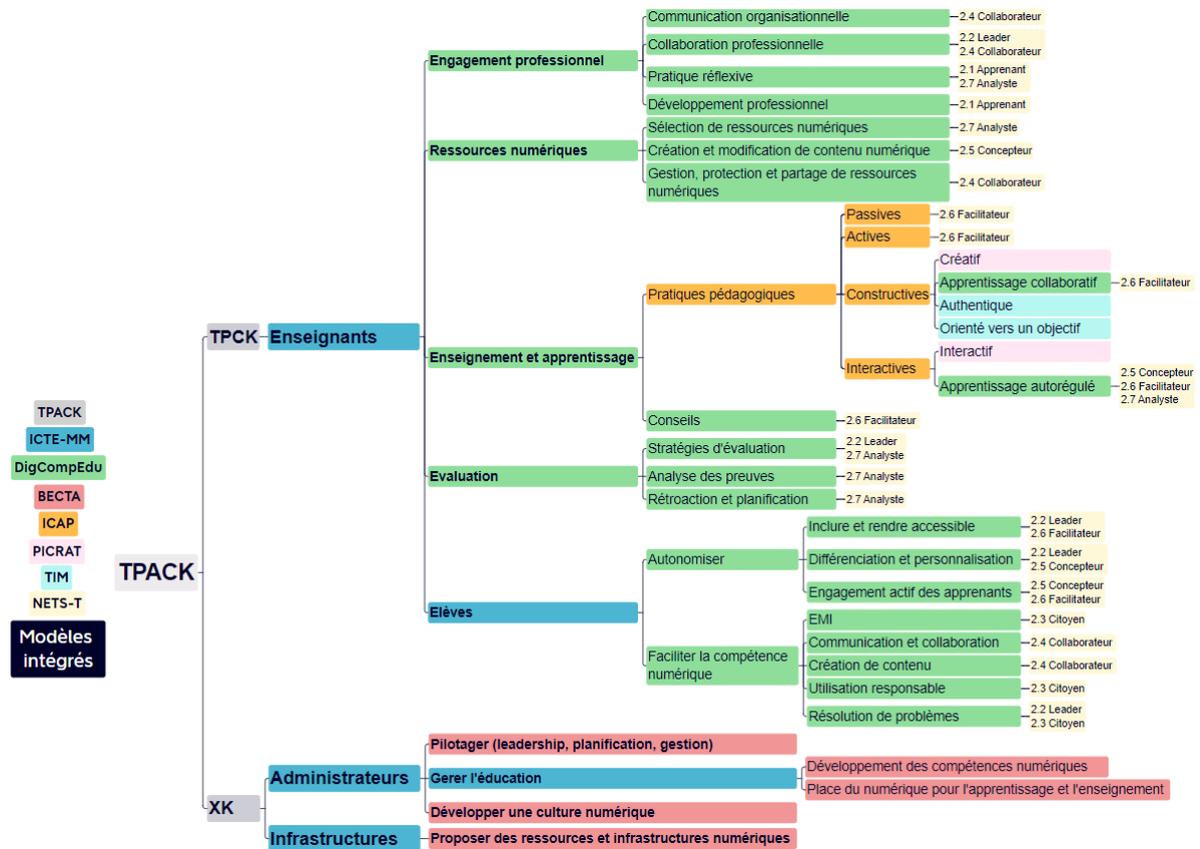


Figure 8 : (a) Modèles intégrés et (b) Critères de caractérisation sélectionnés

Le modèle unifié proposé comporte 3 domaines principaux (issus de l'ICTE-MM) : Enseignants, Administrateurs et Infrastructure. Le domaine Enseignants 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.

5.2 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 les modèles CIT, TPACK, 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.

Tableau 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 Développement	1 Initial	5
BECTA	5 Maturité	4 Avancé	3 Compétent		2 Autonomisation		1 Pas de maturité	5
DigComp Edu	Pionnier (C2)	Leader (C1)	Expert (B2)	Intégrateur (B1)	Explorateur (A2)	Nouvel arrivant (A1)		6
LoTi	Perfectionnement	Expansion	Intégration	Infusion	Exploration	Sensibilisation	Non-utilisation	7
ICAP			Interactif	Collaboratif	Actif	Passif		4
PICRAT	Transformation		Amplification			Remplacement		3
TIM	Transformation		Infusion	Adaptation	Adoption	Entrée		5
SAMR	Redéfinition	Modification			Augmentation	Substitution		4
CIT Model	Explorateur	Autonomisation				Négociation	Choc	4
TPACK	TPACK						TK, PK, CK	1
NETS-T							Pas de niveau	
Synthèse	Transformation	Développement	Intégration		Amélioration	Substitution	Non-utilisation	6
	Pionnier	Leader	Expert		Explorateur	Nouvel arrivant	Non-utilisateur	6

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 (2003), on peut constater que les niveaux 6, 5, 3, 2, 1 (« Innovateur », « Adoptant précoce », « Majorité acquise », « 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 maîtrise 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 et (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 au 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.

Nous recommandons d'utiliser un modèle en 6 niveaux qui considère soit les processus caractéristiques (Transformation, Développement, Intégration, Amélioration, Substitution, 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 9), à titre indicatif. Des études empiriques doivent être menées pour en définir la forme précise. La classification de Rogers (2003) y a aussi été présentée à titre de comparaison.

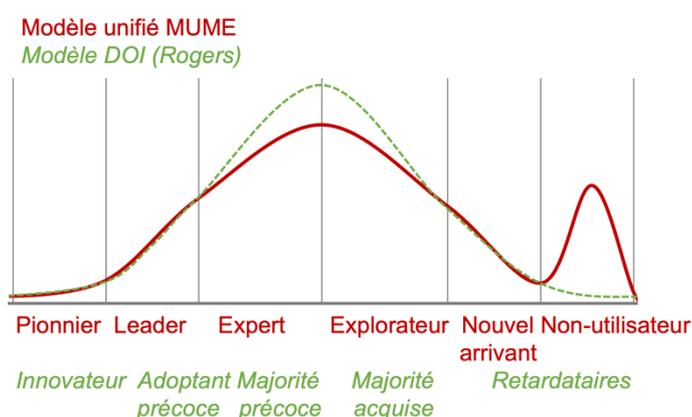


Figure 9 : Comparaison des courbes de diffusion des technologies pour l'éducation pour le DOI de Rogers (2003) et le modèle unifié MUME.

6. VERS UN DIAGNOSTIC UNIFIÉ DE LA MATURITÉ NUMÉRIQUE

Nos travaux sur l'analyse de la maturité numérique se sont poursuivis par l'étude des moyens de mesure de cette maturité. En effet, la plupart des modèles intègrent un outil de diagnostic permettant d'évaluer la maturité numérique des individus. Notre objectif était double : comparer les différents moyens pour vérifier leur cohérence et construire un outil de ce type pour faire des diagnostics adaptés pour le modèle MUME.

6.1 LES OUTILS DE DIAGNOSTICS SOURCES

La revue de littérature a permis d'identifier et sélectionner trois outils évaluatifs utilisés par certains modèles constitutifs de MUME : à savoir l'ICAP-TS (Antonietti *et al.*, 2023), le TPACK.x (Dorsaz, 2022 ; Schmid *et al.*, 2020), et le SELFIE (Costa *et al.*, 2021). Nous avons aussi utilisé le questionnaire TNE 95-02 développé dans une précédente recherche (Michel et Pierrot, 2022).

Le TPACK.x correspond à un questionnaire d'auto-évaluation comprenant 36 items en fonction des 8 domaines du TPACK, dont celui correspondant au contexte. Le TPACK.x est

une traduction francophone validée (Dorsaz, 2022) du questionnaire existant TPACK.xs (Schmid *et al.*, 2020) et qui utilise une échelle d'accord en 5 niveaux (de 1 « fortement en désaccord à 5 « tout à fait d'accord »). Nous n'avons retenu que les items correspondant au domaine de connaissances XK (le contexte) répartis en 3 sous-domaines : XK_{ma} pour les connaissances contextuelles au niveau macro, XK_{me} pour celles au niveau méso et XK_{mi} pour celles au niveau micro. Le niveau macro vise à appréhender les principales dimensions pertinentes dans l'intégration du numérique et concerne les actions politiques, culturelles, économiques et stratégiques relatives au déploiement des technologies. Le niveau micro se situe à l'échelle de la classe où se déploie l'activité d'enseignement, il tient compte des caractéristiques des apprenants et des enseignants. Entre ces deux niveaux, le niveau méso décrit l'environnement qui encadre la classe (prise en compte des pratiques pédagogiques au sein de l'établissement ou des pratiques des familles par exemple).

Le questionnaire d'auto-évaluation ICAP-TS (Antonietti *et al.*, 2023) est décomposé en 4 sous-échelles en fonction de la taxonomie ICAP pour évaluer des activités d'apprentissage de type passif, actif, constructif ou interactif, soutenues par le numérique. La sous-échelle « passif » concerne la mise en œuvre d'activités de type transmissives avec le numérique. La sous-échelle « actif » vise les activités qui s'appuie sur l'utilisation par les élèves du numérique pour appliquer des connaissances. La sous-échelle « constructif » décrit des activités d'apprentissage avec le numérique où les élèves construisent de nouvelles connaissances. La sous-échelle « interactif » concerne la mise en œuvre d'activités avec le numérique de type collaboratives. En complément des 4 sous-échelles, le questionnaire ICAP-TS inclut des questions relatives à la fréquence d'utilisation de 12 outils numériques (logiciel de traitement de texte ou service de test en ligne par exemple). Le questionnaire s'appuie sur une échelle de fréquence en 5 points (0 - Presque jamais, 4 - Presque que tous les jours). Le questionnaire est proposé en anglais, il a fait l'objet d'une traduction.

L'outil d'auto-évaluation SELFIE (Costa *et al.*, 2021) est adapté au DigCompEdu et inclut trois questionnaires, dont un pour les enseignants. Ce questionnaire, multilingue, est composé de plusieurs éléments pour collecter les attitudes et croyances des répondants, leurs pratiques et compétences numériques ainsi que des informations sur les ressources mises à leur disposition et leur contexte d'enseignement. Pour l'essentiel des questions, le questionnaire repose sur une échelle de fréquence en 5 points (1 - pas du tout d'accord à 5 - tout à fait d'accord). Les items relatifs à l'auto-évaluation de la maîtrise des compétences numériques sont calqués sur le référentiel de compétences DigCompEdu, avec une échelle en 7 points (A0 - je ne sais pas comment faire à C2 - je conçois de nouvelles manières de le faire).

Enfin, le questionnaire TNE 95-02 (Michel et Pierrot, 2022) est organisé en 4 parties pour identifier le contexte professionnel d'enseignement des répondants, leurs pratiques numériques avant, pendant et après la période de crise sanitaire, leur expérience d'enseignement durant le confinement et leurs caractéristiques personnelles. Ce questionnaire présente l'intérêt d'être adapté au contexte d'enseignement français, dans la mesure où il a été conçu pour être diffusé auprès d'enseignants des 1^{er} et 2nd degrés impliqués dans le programme Territoire numérique éducatif.

6.2 LE DÉVELOPPEMENT D'UN NOUVEAU QUESTIONNAIRE

Le développement du nouveau questionnaire pour mesurer les niveaux de maturité a consisté à identifier les groupes de questions mobilisables dans les outils d'évaluation décrits précédemment, et cela de façon complémentaire. La fusion de l'ensemble des outils a permis

de produire une première version du questionnaire incluant 23 questions regroupées en 3 parties :

- caractéristiques personnelles (14 questions à choix unique ou multiple, provenant du questionnaire TNE 95-02) ;
- utilisation du numérique (8 questions de type échelle de fréquence ou d'accord, provenant du TPACK.x et ICAP-TS) ;
- compétences numériques (1 question de type échelle en 7 points provenant du SELFIE).

À ce stade du développement du questionnaire, les questions du TPACK.x, du SELFIE et de TNE 95-02 étant déjà disponibles en français, seuls les items de l'échelle ICAP-Ts ont fait l'objet d'une première traduction et adaptation culturelle. Les items ont d'abord été traduits de l'anglais au français puis adaptés de sorte à être applicables au contexte français.

Le questionnaire élaboré a ensuite fait l'objet d'un prétest qualitatif. La compréhensibilité et la pertinence des questions ont été discutées, entre les mois de mars et d'avril 2023, avec 6 enseignants-chercheurs disposant tous d'une expérience solide dans l'accompagnement au déploiement de dispositifs numériques dans les 1^{er} et 2nd degrés. Ce prétest a permis de mettre en évidence la complexité associée à la compréhension de l'échelle TPACK.x. Plus spécifiquement, les questions relatives à la dimension macro (XK_{ma}) ont été perçues par l'ensemble des experts comme inadaptées au contexte socioculturel français (par exemple les questions du type « *Je suis au courant de la culture régionale et nationale en matière d'utilisation des technologies pour l'enseignement* »). Le deuxième retour a porté sur la variété et le manque d'homogénéité des échelles mobilisées, celle du TPACKx et ICAP-TS étant en 5 points avec un premier niveau à 0 ou à 1, celles du SELFIE en 7 points. Le troisième retour porte sur la traduction de certains items de l'échelle ICAP-TS, jugée inadaptée. C'est le cas par exemple du libellé « *pour qu'ils construisent individuellement de nouvelles connaissances* », jugé trop complexe. Enfin, dans l'ensemble, les experts ont trouvé certaines questions des parties A et B redondantes et le questionnaire trop long pour être acceptable.

Le questionnaire a été retravaillé pour simplifier la formulation de certaines questions sur la base des retours des experts. Nous avons néanmoins conservé certaines redondances dans les questions de manière à pouvoir comparer et évaluer la cohérence des résultats selon les quatre questionnaires sources.

La nouvelle version du questionnaire inclut toujours 3 parties :

- caractéristiques personnelles (7 questions à choix unique ou multiple, provenant du questionnaire TNE 95-02) ;
- utilisation du numérique (7 questions de type échelle de fréquence ou d'accord en 7 points, provenant du TPACK.x et ICAP-TS) ;
- compétences numériques (1 question de type échelle en 7 points provenant du SELFIE).

6.3 ÉVALUATION PRÉLIMINAIRE DU QUESTIONNAIRE

Dans l'objectif de vérifier sa validité, le nouveau questionnaire a été diffusé au mois de juin 2023 dans le cadre du projet CoAI-Datastim auprès d'enseignants de l'académie de

Paris des 1^{er} et 2nd degrés via l'ENT PCN⁵ opéré par l'entreprise EDIFICE⁶. L'intérêt de cette collecte de données est qu'en complément du questionnaire, nous disposons, par EDIFICE, des traces d'activité sur les ENT PCN. Nous pourrions ainsi comparer et mettre en perspective les usages déclarés concernant l'ENT dans le questionnaire avec les usages effectifs de l'ENT. L'analyse des données est en cours. Nous proposons ici quelques résultats préliminaires.

L'échantillon analysé comprend 143 participants, répartis entre 86 enseignants du 1^{er} degré, 18 participants du 2nd degré, et 39 participants sans réponse. Les femmes constituent la majorité (101), comparées aux 22 enseignants hommes et aux 20 participants sans réponse. Dans l'échantillon, les enseignants de 41 à 50 ans sont les plus nombreux (52), suivis par ceux de plus de 51 ans (43). Les tranches d'âge inférieures à 30 ans et de 31 à 40 ans comptent respectivement 6 et 23 enseignants. Treize participants n'ont pas complété cette question. La majorité des enseignants (88) ont plus de 10 ans d'expérience, tandis que 23 en ont entre 3 et 10 ans. Cinq enseignants ont moins de 3 ans d'expérience, et 8 enseignants ont une expérience qualifiée comme « Autre ». Les données d'expérience ne sont pas disponibles pour 19 enseignants.

La validité du questionnaire a été évaluée en s'appuyant sur la mesure de la fiabilité des questions, à travers le coefficient α de Cronbach. La valeur de ce coefficient s'établit généralement entre 0 et 1 pour des sous-échelles ou groupes de questions et est considérée comme *acceptable* à partir de 0,70. L'ensemble des questions de notre questionnaire obtient un score supérieur à cette valeur seuil (voir tableau 3), mais il convient de préciser que notre échantillon est ici restreint : il est préconisé de collecter des données auprès d'un échantillon de 200 à 300 participants pour appréhender l'hétérogénéité des réponses.

Tableau 3 : α de Cronbach des groupes de questions relatives aux différents thèmes du questionnaire

Groupes de variables	α de Cronbach
Utilisation du numérique en classe	0,959
Utilisation du numérique par les élèves (ICAP-TS)	0,945
Utilisation du numérique par les enseignants (ICAP-TS)	0,922
Utilisation du numérique et prise en compte du contexte de la classe XKmi (TPACK.x)	0,906
Utilisation du numérique et prise en compte du contexte de l'environnement XKme (TPACK.x)	0,709
Compétences numériques (SELFIE)	0,967

L'analyse préliminaire des résultats révèle en outre des indices sur les dynamiques d'intégration du numérique dans les pratiques pédagogiques des enseignants et plus globalement leur montée en maturité. Comme nous l'avons fait précédemment (Michel *et al.*, 2021 ; Michel et Pierrot, 2022), nous avons identifié des classes de maturité des répondants selon leur auto-évaluation des compétences numériques, à l'aide d'une classification K-Means en 7 niveaux. De manière à comparer les mesures de chaque modèle,

⁵ <https://www.parisclassenumerique.fr/>

⁶ <https://edifice.io>

nous avons opéré cette classification sur chaque sous-groupe de questions liées aux questionnaires sources, par exemple CDig pour les questions liées au SELFIE ou CEnclasse pour les questions de l'ICAP-TS. L'analyse comparée des différentes classifications est en cours, nous la présenterons ultérieurement. À titre préliminaire nous présentons dans le paragraphe suivant les résultats de la classification réalisée sur la base des questions extraites du SELFIE.

Les 7 classes, CDig1 à CDig7, listées dans la première ligne de le tableau 4, permettent d'identifier des dynamiques selon les usages et compétences décrits dans la première colonne du tableau. Ces usages et compétences vont de 0 (je ne sais pas comment faire) à 4 (je le fais régulièrement et je conçois de nouveaux moyens pour le faire), ils sont classés selon un niveau d'usage global (dernière colonne). La première classe d'enseignants (CDig1, 37 enseignants) correspond à ceux qui déclarent le moins savoir utiliser le numérique, qui sert essentiellement des objectifs de communication (rattachés au domaine 1 du DigCompEdu). Les classes CDig2 à 5 (respectivement, 1, 17, 8 et 2 enseignants) ont une utilisation plus diversifiée et régulière du numérique que la CDig1. Leurs usages et compétences s'inscrivent autour des domaines 2 (ressources numériques), 6 (développement des compétences numériques des apprenants) et 3 (enseignement et apprentissage). La progression d'une classe à l'autre est guidée par les items réflexifs et d'autoformation. La classe CDig6, 2 enseignants, et la classe CDig7, 4, correspondent aux enseignants dont les usages sont les plus développés, y compris dans les domaines du DigCompEdu d'évaluation et d'autonomisation des apprenants.

Tableau 4 : Classes d'enseignants selon leur maîtrise des compétences numériques

Domaine	Classe DigComp Ens/ Usages	CDig1	CDig2	CDig3	CDig4	CDig5	CDig6	CDig7	Niveau d'usage
D4	[Utiliser les outils numériques pour évaluer l'apprentissage de manière formative et sommative]	0,01	0,00	0,24	0,58	0,50	0,70	2,05	0,58
D5	[Utiliser les outils numériques pour s'adapter aux différents niveaux et rythmes d'apprentissage des élèves]	0,01	0,00	0,73	0,10	0,10	0,90	2,75	0,65
D2	[Organiser les contenus numériques pour qu'ils soient faciles à utiliser pour les élèves, les parents et les enseignants]	0,05	0,00	0,91	0,68	0,10	0,50	2,50	0,68
D3	[Utiliser des technologies innovantes (Robotique, IA, RV) pour explorer de nouveaux contenus et expériences d'apprentissage]	0,01	0,00	0,12	0,53	0,10	2,40	2,10	0,75
D3	[Utiliser les outils numériques pour aider les élèves à résoudre des problèmes et mieux comprendre]	0,08	0,00	0,35	0,10	0,10	3,00	2,75	0,91
D3	[Utiliser les outils numériques pour aider les élèves à apprendre de manière autonome]	0,05	0,00	0,46	0,20	0,10	3,00	2,75	0,94
D4	[Utiliser les outils numériques pour suivre les progrès individuels des élèves]	0,01	0,00	0,29	0,63	0,50	2,40	2,75	0,94
D5	[Tenir compte des facteurs affectifs des élèves (motivation, fatigue, etc.) pour garantir un environnement d'apprentissage favorable]	0,00	0,00	0,72	1,13	1,00	1,50	2,25	0,94
D3	[Utiliser les outils numériques pour que les élèves travaillent ensemble et communiquent]	0,06	0,00	0,35	0,08	0,10	4,00	2,50	1,01
D6	[Donner aux élèves les moyens d'utiliser les outils numériques en toute sécurité (gestion des risques liés à leur bien-être physique, psychologique et social)]	0,01	0,00	0,86	0,35	0,10	3,50	2,50	1,04
D5	[Tenir compte des conditions d'accès aux outils numériques (disponibilité matérielle, compétences, etc.) pour garantir un environnement d'apprentissage favorable]	0,00	0,00	0,60	0,50	1,00	2,50	2,75	1,05
D3	[Utiliser les outils numériques pour que les élèves produisent des contenus et s'expriment]	0,06	0,00	0,35	0,20	0,30	3,50	3,00	1,06
D6	[Donner aux élèves les moyens d'utiliser les outils numériques de manière responsable et éthique (gérer leur identité numérique, empreinte numérique, réputation numérique, etc.)]	0,01	0,00	0,68	0,13	0,10	4,00	2,75	1,09
D5	[Tenir compte des caractéristiques socio-démographiques des élèves (âge, milieu socio-économique, culture, langue maternelle, etc.) pour garantir un environnement d'apprentissage favorable]	0,02	0,00	0,71	0,88	1,00	2,50	2,75	1,12
D2	[Utiliser les outils numériques pour créer ou modifier des ressources numériques pour mes cours]	0,14	0,00	1,02	0,20	3,50	0,90	2,75	1,22
D2	[Utiliser les outils numériques pour trouver et sélectionner des ressources numériques pour mes cours]	0,24	0,00	1,00	0,30	3,50	1,90	2,25	1,31
D3	[Utiliser les outils numériques pour aider les élèves à apprendre plus efficacement]	0,29	0,00	0,89	0,73	0,80	4,00	2,50	1,32
D1	[Utiliser les outils numériques pour aider les élèves et les enseignants à réfléchir sur leur travail]	0,01	3,00	0,44	0,15	0,10	4,00	2,50	1,46
D2	[Respecter les lois sur la propriété intellectuelle lors de l'utilisation des outils numériques]	0,05	0,00	1,62	1,48	2,50	3,50	2,05	1,60
D1	[Utiliser les outils numériques pour travailler avec mes collègues et d'autres personnes impliquées dans l'éducation]	0,39	3,00	1,39	2,25	1,90	0,70	3,00	1,80
D1	[Utiliser les outils numériques pour apprendre de nouvelles choses pour mon travail]	0,16	0,00	1,27	0,90	3,50	4,00	3,00	1,83
D2	[Protéger les informations personnelles des élèves, des parents et des enseignants en ligne]	0,10	3,00	0,96	2,25	2,50	3,00	2,75	2,08
D1	[Réfléchir à la manière dont j'utilise les outils numériques dans mon travail]	0,09	4,00	0,80	2,10	2,50	3,00	2,50	2,14
D1	[Utiliser les outils numériques dans mon école pour communiquer avec les élèves, les collègues et les parents]	0,66	3,00	1,56	2,48	2,40	1,90	3,00	2,14
	Nb Ens	37,00	1,00	17,00	8,00	2,00	2,00	4,00	
	Variance intra-classe	2,45	0,00	10,02	23,57	17,74	28,18	16,41	
	Moyenne	0,10	0,67	0,76	0,79	1,18	2,55	2,60	

7. CONCLUSION 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 considérant 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. Les questionnaires identifiés grâce à la revue de littérature présentent l'intérêt majeur d'offrir une vue des niveaux de maturité des enseignants. Cependant, ces moyens sont tous basés sur des données auto-rapportées et correspondent à un diagnostic à un temps donné. De manière à pouvoir comprendre la progression des usages (Michel et Pierrot, 2022), une approche mixte d'observation et d'analyse des niveaux de maturité numérique des enseignants paraît pertinente. De cette manière, ce travail se poursuit actuellement par le développement d'un outil de diagnostic de la maturité basé sur un questionnaire. À ce stade, il nous reste à poursuivre l'analyse des données pour en vérifier la validité. Plus globalement, cette mise en application du modèle MUME, orientée vers la collecte de données déclaratives, sera combinée à d'autres types de données, collectées automatiquement (logs d'usage des outils numériques) ou non (enregistrements vidéo d'activités d'enseignement). Notre objectif est, à terme, de nous appuyer sur les potentialités offertes par les travaux des *Learning* et *Teaching Analytics* pour proposer un tableau de bord de la maturité numérique à destination des enseignants.

8. REMERCIEMENTS

Ce travail a été réalisé en collaboration avec l'entreprise *Édifice* et financé dans le cadre du projet *CoAI-DataStim* (Académie de Paris).

9. BIBLIOGRAPHIE

Antonietti, C., Schmitz, M.-L., Consoli, T., Cattaneo, A., Gonon, P. et Petko, D. (2023). Development and validation of the ICAP Technology Scale to measure how teachers integrate technology into learning activities. *Computers & Education*, 192, 104648. <https://doi.org/10.1016/j.compedu.2022.104648>

- Backfisch, I., Lachner, A., Stürmer, K. et Scheiter, K. (2021). Variability of teachers' technology integration in the classroom: A matter of utility! *Computers & Education*, 166, 104159. <https://doi.org/10.1016/j.compedu.2021.104159>
- BECTA. (2008). *Measuring e-maturity amongst work-based learning providers 2008: final report*. British Educational Communications and Technology Agency (BECTA).
- Blundell, C. N., Mukherjee, M. et Nykvist, S. (2022). A scoping review of the application of the SAMR model in research. *Computers and Education Open*, 3.
- Boell, S. et Cecez-Kecmanovic, D. (2014). A Hermeneutic Approach for Conducting Literature Reviews and Literature Searches. *Communications of the Association for Information Systems*, 34(1). <https://doi.org/10.17705/1CAIS.03412>
- Carvalho, J., Pereira, R. H. et Rocha, Á. (2018, juin). *Maturity models of education information systems and technologies: A systematic literature review*. 13th Iberian Conference on Information Systems and Technologies (CISTI).
- 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. et Yaghmourian, D. L. (2018). Translating the ICAP Theory of Cognitive Engagement Into Practice. *Cognitive Science*, 42(6), 1777-1832.
- Costa, P., Castaño-Muñoz, J. et Kamylylis, P. (2021). Capturing schools' digital capacity: Psychometric analyses of the SELFIE self-reflection tool. *Computers & Education*, 162, 104080. <https://doi.org/10.1016/j.compedu.2020.104080>
- Crompton, H. et Sykora, C. (2021). Developing instructional technology standards for educators: A design-based research study. *Computers and Education Open*, 2.
- Dorsaz, D. (2022). *Le questionnaire TPACK-X. Développement et validation d'un outil d'auto-positionnement mesurant les connaissances TPACK et contextuelles des formateur-trice-s en Suisse Romande* [mémoire de maîtrise, Université de Genève]. <https://tecfa.unige.ch/tecfa/maltt/memoire/Dorsaz2022.pdf>.
- Franklin, C. et Bolick, C. (2007). *Technology Integration: A Review of the Literature* (p. 1482-1488). Association for the Advancement of Computing in Education (AACE).
- Harrison, C., Tomás, C. et Crook, C. (2014). An e-maturity analysis explains intention-behavior disjunctions in technology adoption in UK schools. *Computers in Human Behavior*, 34, 345-351.
- ISTE (2017). ISTE Standards for Educators. *International Society for Technology in Education*. <https://www.iste.org/standards/iste-standards-for-teachers>
- Jabłonowska, M. et Wiśniewska, J. (2021). *Europejskie ramy kompetencji cyfrowych nauczycieli - kluczowe obszary badania poziomu umiejętności i ich implikacje*. <https://doi.org/10.34866/NB0C-2E79>
- Kimmons, R., Graham, C. R. et West, R. E. (2020). The PICRAT Model for Technology Integration in Teacher Preparation. *Contemporary Issues in Technology and Teacher Education*, 20(1), 176-198.
- Kozdras, D. et Welsh, J. (2018). Enter the Matrix: A Pedagogy for Infusing Technology. Dans *Society for Information Technology & Teacher Education International Conference* (p. 536-541). Association for the Advancement of Computing in Education (AACE).

- Laaber, F., Florack, A., Koch, T. et Hubert, M. (2023). Digital maturity: Development and validation of the Digital Maturity Inventory (DIMI). *Computers in Human Behavior*, 143, 107709. <https://doi.org/10.1016/j.chb.2023.107709>
- Leite, L. et Lagstedt, A. (2021). The Collective Integration of Technology (CIT) Model: Helping Teachers Incorporate Technology Meaningfully in their Everyday Work. *The International Journal of Education and Development using Information and Communication Technology*, 17, 249-268.
- McCarthy, A., Maor, D., McConney, A. et Cavanaugh, C. (2023). Digital transformation in education: Critical components for leaders of system change. *Social Sciences & Humanities Open*, 8, 100479. <https://doi.org/10.1016/j.ssaho.2023.100479>
- Michel, C. et Pierrot, L. (2022). Towards Modelling the Technology Integration in Elementary School. A Diachronic Study of Teachers' Digital Practices During and After Covid-19 Lockdown. Dans I. Hilliger, P. J. Muñoz-Merino, T. De Laet, A. Ortega-Arranz et T. Farrell (dir.), *Educating for a New Future: Making Sense of Technology-Enhanced Learning Adoption* (vol. 13450, p. 201-214). Springer International Publishing. https://doi.org/10.1007/978-3-031-16290-9_15
- Michel, C. et Pierrot, L. (2023). *Les modèles de la maturité numérique des enseignants Lot 1 - État de l'art* (Projet Covalence 1 (DataStim)).
- Michel, C., Pierrot, L. et Solari-Landa, M. (2021). VLE Limits and Perspectives for Digital Integration in Teaching Practices. Dans T. De Laet, R. Klemke, C. Alario-Hoyos, I. Hilliger et A. Ortega-Arranz (dir.), *Technology-Enhanced Learning for a Free, Safe, and Sustainable World* (p. 96-109). Springer International Publishing. https://doi.org/10.1007/978-3-030-86436-1_8
- Mishra, P. (2019). Considering Contextual Knowledge: The TPACK Diagram Gets an Upgrade. *Journal of Digital Learning in Teacher Education*, 35(2), 76-78.
- Mishra, P. et Koehler, M. J. (2006). Technological Pedagogical Content Knowledge: A Framework for Teacher Knowledge. *Teachers College Record*, 108(6), 1017-1054.
- Moersch, C. (1995). Levels of Technology Implementation (LoTi): A framework for measuring classroom technology use. *Learning and leading with technology*, 23(3).
- Pee, L. G. et Kankanhalli, A. (2009). A Model of Organisational Knowledge Management Maturity Based on People, Process, and Technology. *Journal of Information & Knowledge Management*, 08, 79-99.
- Plantard, P. et Serreau, M. (2024). Le numérique comme fait social total. *Canadian Journal of Learning and Technology*, 49, 1-19. <https://doi.org/10.21432/cjlt28463>
- Puentedura, R. (2012). *SAMR - A research perspective*. <http://www.hippasus.com/trpweblog/archives/2014/11/28/SAMRLearningAssessment.pdf>
- Redecker, C. (2017). *European Framework for the Digital Competence of Educators: DigCompEdu*. Punie, Y. (ed). EUR 28775 EN. Publications Office of the European Union.
- Ristić, M. (2018). E-Maturity in Schools. *Croatian Journal of Education*, 19(3), 317-334.
- Rogers, E. M. (2003). *Diffusion of Innovations* (5^e éd.). Free Press.

- Sackstein, S., Matthee, M. et Weilbach, L. (2022). Theories and Models Employed to Understand the Use of Technology in Education: A Hermeneutic Literature Review. *Education and Information Technologies*.
- Schmid, M., Brianza, E. et Petko, D. (2020). Developing a short assessment instrument for Technological Pedagogical Content Knowledge (TPACK.xs) and comparing the factor structure of an integrative and a transformative model. *Computers & Education*, 157, 103967. <https://doi.org/10.1016/j.compedu.2020.103967>
- Solar, M., Sabattin, J. et Parada, V. (2013). A Maturity Model for Assessing the use of ICT in School Education. *Educational Technology & Society*, 16, 206-218.
- Stoltzfus, J. (2006). *Determining Educational Technology and Instructional Learning Skill Sets (DETAILS): A New Approach to the LoTi Framework for the 21st Century*.
- Teichert, R. (2019). Digital Transformation Maturity: A Systematic Review of Literature. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis*, 67(6), 1673-1687.
- Tomczyk, Ł. et Fedeli, L. (2021). *Digital Literacy among Teachers -Mapping Theoretical Frameworks: TPACK, DigCompEdu, UNESCO, NETS-T, DigiLit Leicester* (p. 244-252).
- Underwood, J., Baguley, T., Banyard, P., Dillon, G. et Farrington-Flint, L. (2010). *Understanding the Impact of Technology: Learner and School level factors 2010*.
- Underwood, J. et Dillon, G. (2004). Capturing complexity through maturity modelling. *Technology, Pedagogy and Education*, 13(2), 213-225.

Discriminations intersectionnelles : approfondir l'évaluation de l'équité algorithmique. L'exemple de la prédiction de la réussite académique avec des données issues de cours en ligne

Intersectionality : deepen algorithmic fairness evaluation. The case study of academic performance prediction using data from online courses

Méline VERGER¹; François BOUCHET¹; Sébastien LALLÉ¹; Vanda LUENGO¹

¹*Sorbonne Université, CNRS, LIP6, F-75005 Paris, France*

Résumé. Évaluer l'équité algorithmique des modèles prédictifs utilisés en éducation est devenu crucial pour garantir que, déployés, ils ne soient pas biaisés envers certains apprenants. Jusqu'à présent, les analyses se sont concentrées sur l'évaluation de l'équité algorithmique vis-à-vis d'attributs sensibles, comme le genre, présents dans les données, indépendamment les uns des autres. Or, la théorie de l'intersectionnalité de Crenshaw (1989) défend l'idée que les influences conjointes de plusieurs attributs sensibles produisent des discriminations uniques et différentes pour certaines sous-groupes d'individus. Ainsi, nous proposons dans cet article d'étendre les travaux de Verger, Bouchet *et al.* (2023) avec des analyses supplémentaires sur les discriminations intersectionnelles présentes dans les prédictions des modèles. Ces modèles ont été utilisés dans le cadre de la prédiction de la réussite à des cours en ligne, au moyen de données éducatives ouvertes, plus précisément les données du corpus OULAD (Kuzilek *et al.*, 2017). Nos résultats ont permis de mettre en lumière des discriminations algorithmiques qui n'étaient pas identifiables à partir des analyses classiques ainsi que de déterminer l'influence de chaque attribut sur les discriminations grâce à leurs interactions avec les autres attributs.

Mots-clés : équité algorithmique, intersectionnalité, prédiction

Abstract. *Assessing the algorithmic fairness of predictive models used in education has become crucial to ensure that, when deployed, they are not biased in favor or against certain learners. Until now, analyses have focused on assessing algorithmic fairness with regard to sensitive attributes in the data, such as gender, independently of each other. However, Crenshaw (1989)'s theory of intersectionality defends the idea that the influences of several sensitive attributes together produce unique and different discriminations for certain sub-groups of individuals. Thus, in this paper, we propose to extend (Verger, Bouchet *et al.*, 2023)'s work with additional analyses of intersectional discriminations that are in the outcomes of predictive models. These models were used in the context of predicting success in online courses, using open educational data, specifically data from OULAD (Kuzilek *et al.*, 2017). Our results shed light on algorithmic discriminations that were not identifiable from traditional analyses, as well as determining the influence of each attribute on discriminations through their interactions with other attributes.*

Keywords : *algorithmic fairness, intersectionality, prediction*

1. INTRODUCTION

Dans cette section, nous examinons pourquoi et comment nous nous intéressons aux discriminations intersectionnelles. Nous présentons d’abord le vocabulaire qui sera employé tout au long de l’article (Section 1.1). Puis, nous expliquons le contexte de notre recherche (Section 1.2), suivi de ses motivations et questions de recherche (Section 1.3). Enfin, nous décrivons les contributions attendues (méthodes et résultats principaux) (Section 1.4) ainsi que le plan de l’article (Section 1.5).

1.1. TERMINOLOGIE

Avant d’entrer dans les détails, nous précisons quelques éléments terminologiques. Premièrement, l’*équité algorithmique*, qui est en fait une traduction des termes anglais plus largement utilisés *algorithmic fairness*, fait référence non pas à une équité mais à une égalité de traitement des individus par les systèmes composés d’au moins un modèle d’apprentissage automatique. L’égalité de traitement doit être faite, indépendamment de caractéristiques telles que le sexe ou l’origine ethnique¹ (Barocas *et al.*, 2019). Par exemple, un modèle qui prédit l’admission à l’université doit produire des résultats indépendants du genre de l’apprenant. À l’inverse, nous parlerons de *discriminations algorithmiques* lorsqu’un système s’avèrera ne pas produire cette égalité de traitement en fonction de l’une ou l’autre des caractéristiques en question. Dans le contexte de cet article, *équité* et *discrimination* seront parfois utilisés sans l’adjectif *algorithmique* par souci de simplicité.

Deuxièmement, ces caractéristiques, auxquelles nous ne voulons pas que les résultats d’un système accordent un traitement discriminatoire, sont appelées *attributs sensibles* (ou en anglais *protected attributes* ou *sensitive features* (Kizilcec et Lee, 2022; Pessach et Shmueli, 2022)). Ces attributs peuvent correspondre à des données dites sensibles en tant que données personnelles, telles que des données socio-démographiques (comme les deux évoquées précédemment), mais nous emploierons ces termes d’*attributs sensibles* de manière plus générale pour désigner les attributs vis-à-vis desquels nous ferons des analyses d’équité.

1.2. CONTEXTE

L’équité algorithmique, c’est-à-dire l’équité dans les systèmes utilisant de l’apprentissage automatique, combinant des modèles mathématiques avec des données, tels que les algorithmes de prédictions, a acquis une importance cruciale en raison de l’utilisation croissante de ces systèmes informatiques. Un exemple d’un tel système en éducation est celui des systèmes adaptatifs, qui utilisent des algorithmes de prédiction pour prédire si l’élève va réussir ou non une tâche, un exercice, un cours, un module, ce qui permet, entre autres, d’adapter l’exercice, la tâche, la recommandation de ressources, le cours ou le parcours.

De la même manière qu’il est courant de tester plusieurs algorithmes de prédiction sur un même corpus de données, car ne donnant pas les mêmes résultats, il est également possible que ces algorithmes donnent des résultats différents vis-à-vis des groupes d’individus. Aussi, l’algorithme le plus performant en termes de prédiction n’est pas nécessairement l’algorithme le plus *équitable* au regard de différents groupes d’individus. Si l’évaluation de la performance prédictive est classique dans ce type de systèmes, l’évaluation de l’équité algorithmique est quant à elle plus récente et est devenue nécessaire afin de garantir que les

1. 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.

systèmes déployés ne soient pas biaisés en faveur ou au détriment de certains groupes d'individus (Buolamwini et Gebru, 2018; Gardner *et al.*, 2019; Lee et Kizilcec, 2020; Mehrabi *et al.*, 2022; Verger, 2022). Ainsi, l'équité algorithmique constitue un enjeu éthique, mais aussi un impératif sociétal, émanant notamment de directives et réglementations de plusieurs instances².

Les discriminations algorithmiques peuvent apparaître à différentes phases de développement des systèmes informatiques, passant par la collecte des données, leur pré-traitement, l'entraînement du modèle et enfin leur usage (Verger, 2022). Les causes de ces discriminations se retrouvent donc soit dans les données, les algorithmes ou les interventions humaines suite aux résultats. En ce qui concerne les sources de discrimination liées aux données, elles peuvent être dues aux discriminations dites réelles ou historiques (c'est-à-dire observables depuis longtemps dans une société, et se reflétant dans les données), ou bien elles peuvent être relatives à une représentation, une collecte ou un calibrage inadéquates des données. C'est donc dans la phase de collecte et pré-traitement que ces discriminations peuvent être étudiées. Ici, nous cherchons à évaluer l'équité algorithmique en sortie d'entraînement du modèle, avant son usage. Autrement dit, nous allons analyser les discriminations algorithmiques présentes dans les résultats du modèle.

1.3. MOTIVATIONS ET QUESTIONS DE RECHERCHE

Dans un travail antérieur (Verger, Bouchet *et al.*, 2023), nous avons évalué l'équité algorithmique selon certains attributs sensibles (voir Section 1.1), étudiés individuellement. Nous avons notamment combiné les données ouvertes OULAD (Kuzilek *et al.*, 2017) avec des modèles prédictifs très couramment utilisés en éducation (régression logistique, arbre de décision, k plus proches voisins et naïf bayésien). Nous les avons évalués avec une nouvelle mesure d'équité, la MADD (*Model Absolute Density Distance*), que nous avons spécifiquement proposée dans Verger, Lallé *et al.* (2023) pour éviter un écueil courant dans l'évaluation de l'équité algorithmique (voir Section 2). Pour résumer, cette évaluation a donné lieu à une étude numérique et visuelle de l'influence de chaque attribut sensible sur l'équité des résultats des modèles. Ceci nous avait permis de montrer qu'il n'y avait pas de lien direct entre les discriminations observées dans les données et les discriminations obtenues en sortie des modèles. Ainsi, malgré le biais de genre dans les données du cours de sciences sociales (fortes corrélations avec l'attribut genre, forts déséquilibres entre les deux groupes le constituant avec une large majorité de femmes), notre analyse montre que c'est un autre attribut sensible, "pauvreté", qui est à l'origine des discriminations algorithmiques les plus importantes. Cela montre l'importance d'évaluer en profondeur l'équité des modèles produits par ce type d'algorithme en plus des évaluations liées aux autres sources de discrimination.

Ainsi, dans cet article, nous nous proposons d'aller plus loin dans l'évaluation de l'équité algorithmique en considérant non pas l'influence individuelle, mais l'influence simultanée de plusieurs attributs sensibles. En effet, la théorie de l'intersectionnalité de Crenshaw (1989) défend l'idée que les influences de plusieurs attributs ensemble produisent des discriminations uniques et différentes pour certains sous-groupes d'individus. Un exemple souvent rattaché à cette théorie est celui des discriminations expérimentées en tant que femmes noires, qui diffèrent des discriminations entre hommes et femmes ou entre personnes noires et blanches respectivement (Buolamwini et Gebru, 2018). Par conséquent, nous allons examiner les *discriminations intersectionnelles*, c'est-à-dire les discriminations issues de l'in-

2. Par ordre chronologique : 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* au niveau européen.

fluence de plusieurs attributs sensibles en même temps.

Deux questions de recherche se dégagent alors :

QR1 : Comment évaluer l’influence de plusieurs attributs sensibles simultanément ?

QR2 : Découvre-t-on des discriminations algorithmiques supplémentaires quand on considère les individus à l’intersection de plusieurs attributs sensibles ?

1.4. CONTRIBUTIONS

Pour répondre à ces questions, nous suivons une démarche exploratoire et nous proposons de poursuivre les analyses de Verger, Bouchet *et al.* (2023) en évaluant avec la même mesure, MADD, l’équité vis-à-vis des quatre attributs sensibles considérés précédemment mais pris conjointement : le genre, l’âge, le niveau de pauvreté et le handicap. Pour cela, nous nous intéresserons à tous les *groupes intersectionnels* possibles formés par ces quatre attributs (voir Section 6.2). Plus précisément, un *groupe intersectionnel* est défini comme étant un groupe à l’intersection de plusieurs attributs (Gohar et Cheng, 2023), comme par exemple le groupe des « jeunes hommes aisés sans handicap » qui est à l’intersection des attributs âge, genre, niveau de pauvreté et handicap (plus de détails en Section 4).

Les résultats de cet article sont doubles. D’une part, nous proposons une approche d’évaluation pour l’analyse des discriminations intersectionnelles (QR1) et, d’autre part, les analyses permettent de mettre en lumière des discriminations qu’on ne peut voir avec l’analyse des attributs individuels seule et de comprendre de manière plus fine l’influence de chaque attribut sur les discriminations grâce à leurs interactions avec les autres attributs (QR2).

Cet article participe à faire avancer la recherche sur les discriminations algorithmiques et leur évaluation dans les modèles prédictifs en éducation. Comme indiqué plus haut, les discriminations algorithmiques pouvant être ou non différentes des discriminations réelles, ce type d’analyses permet de fournir des éclairages sur les implications de l’utilisation de ces modèles ainsi que sur les discriminations qui pourraient être découvertes. À notre connaissance, cette approche intersectionnelle pour l’évaluation de l’équité algorithmique n’a été faite qu’une fois avec des données éducatives par Zambrano *et al.* (2024), mais sans parvenir à établir de discriminations entre groupes intersectionnels. Notre étude est ainsi la première en éducation à montrer l’utilité de l’approche intersectionnelle pour mettre en lumière des discriminations algorithmiques qui n’auraient pas pu être détectées autrement. Par ailleurs, cette approche est particulièrement destinée aux chercheuses, chercheurs, développeuses et développeurs de modèles prédictifs en éducation et nous mettons à disposition les données, le code documenté³ ainsi que la librairie Python `maddlib`⁴ que nous avons développée pour faciliter ces analyses.

1.5. PLAN DE L’ARTICLE

L’article est organisé comme suit. La Section 2 décrit l’état de l’art de l’évaluation de l’équité algorithmique en éducation spécifiquement pour la tâche de prédiction de la réussite au niveau des cours. La Section 3 présente la mesure d’équité MADD, et la Section 4 comment analyser l’équité avec des groupes intersectionnels. Puis, la Section 5 introduit le cadre expérimental des analyses réalisées en Section 6. Enfin, la Section 7 examine les forces et limites des expériences et des méthodes proposées, avant de conclure en Section 8.

3. <https://github.com/melinaverger/MADD>

4. <https://pypi.org/project/maddlib>

2. ÉTAT DE L'ART

La problématique de cet article portant sur l'évaluation de l'équité algorithmique en considérant l'intersectionnalité, avec comme cas d'étude les systèmes de prédiction de réussite dans des cours en ligne, nous allons d'abord introduire rapidement la notion de réussite à des cours en ligne et les méthodes informatiques de prédiction utilisées (Section 2.1), pour ensuite présenter les travaux sur l'évaluation de l'équité dans ce type de systèmes (Section 2.2), pour finir avec la prise en compte de l'intersectionnalité (Section 2.3).

2.1. LA PRÉDICTION DE LA RÉUSSITE À DES COURS

De manière générale, la capacité à prédire les performances des apprenantes et apprenants dans un cours, ou un parcours, peut permettre d'améliorer les résultats de l'enseignement (Hellas *et al.*, 2018). Comme l'indiquent ces auteurs, des recherches en psychologie et en sciences de l'éducation cherchent à comprendre les facteurs de réussite dite académique depuis au moins un siècle. Les travaux utilisant des algorithmes de prédiction sont quand à eux apparus plus tard, avec l'accès aux données académiques.

Dans cet article, nous nous intéressons aux recherches sur la prédiction de la réussite en s'appuyant sur des données académiques. Ces recherches se concentrent principalement sur la prédiction de la performance à des cours, notamment en termes de réussite ou d'échec, ainsi que sur leur persévérance ou leur abandon dans des contextes d'apprentissage en ligne ou hybride. Comme l'indique la revue systématique d'Hellas *et al.* (2018), la recherche dans ce domaine étudie les caractéristiques (attributs ou facteurs) susceptibles d'être exploitées pour prédire la réussite ainsi que les algorithmes permettant d'améliorer ces prédictions.

Nous pouvons constater que la prédiction de la réussite à partir des données de cours est actuellement une recherche active dans les domaines des learning analytics, fouille de données éducatives et intelligence artificielle pour l'éducation (Romero et Ventura, 2020). En effet, elle constitue, avec la prédiction précoce d'abandon notamment, une des tâches couramment réalisées dans le domaine de l'analyse des données éducatives. C'est pourquoi l'équité algorithmique est étudiée dans le cadre de cette tâche de prédiction (Deho *et al.*, 2022). Bien que ces recherches puissent être réductrices quant aux facteurs influençant la réussite et l'échec, le centre de notre travail ici est l'évaluation de l'équité algorithmique.

Du point de vue des techniques informatiques, d'après différentes revues de littérature (Hellas *et al.*, 2018; Korkmaz et Correia, 2019), prédire la réussite à des cours est le plus souvent représenté par un problème de classification binaire (réussite/échec). Plusieurs types de modèles sont couramment utilisés, tels que les réseaux bayésiens, les arbres de décision, ou la méthode des k plus proches voisins. Dans la continuité de ces pratiques, nous expérimentons notre approche avec quatre types de modèles de classification binaire très courants en éducation (Hellas *et al.*, 2018; Korkmaz et Correia, 2019) et en particulier sur le corpus OULAD (Alhakbani et Alnassar, 2022; Kuzilek *et al.*, 2017). Nous détaillerons ces choix dans la Section 5 dédiée aux expériences. Par ailleurs, le corpus OULAD a été utilisé dans plusieurs travaux de prédiction connexes (réussite/échec, abandon/complétion des cours, etc. voir (Alhakbani et Alnassar, 2022)), mais sans analyse des discriminations algorithmiques jusqu'à ce que nous les initiions dans Verger, Bouchet *et al.* (2023).

2.2. ÉVALUER L'ÉQUITÉ ALGORITHMIQUE

Pour évaluer l'équité algorithmique, il existent trois approches : causale, de similarité et statistique. Pour la première famille, qui traite de l'équité à l'aide de l'approche causale, l'absence d'inéquité est considérée lorsqu'on change l'appartenance d'un individu à un groupe pour un autre (par exemple "homme" pour "femme") et que son résultat reste inchangé

(Kilbertus *et al.*, 2017). Dans l'approche de similarité, l'absence d'inéquité est considérée lorsque deux individus similaires ont obtenu des résultats similaires. Les techniques de cette approche diffèrent selon la distance qu'elles emploient pour mesurer cette notion de similarité. Pour la troisième famille, fondée sur les statistiques, l'absence d'inéquité est considérée lorsqu'un système produit des résultats similaires entre les groupes. Contrairement aux techniques fondées sur la similarité, celles-ci ne se concentrent pas sur des paires d'individus mais sur des groupes, d'où leur autre appellation : « équité de groupe ». Ces techniques consistent donc à quantifier les différences de performance d'un modèle entre les groupes, par exemple grâce à des comparaisons du taux de d'erreur.

Les trois approches ont leurs avantages et inconvénients. Cependant, il s'avère plus difficile d'utiliser l'approche causale car elle requiert de construire des graphes causaux, ce qui est difficile en pratique (Pearl, 2009). De la même façon, l'approche de similarité nécessite une mesure pertinente qui permette de comparer un à un des individus, ce qui est également difficile à mettre en place sans produire d'autres biais. C'est pour cette raison que l'approche statistique est la plus utilisée. Comme évoqué juste avant, pour évaluer l'équité algorithmique avec l'approche statistique, il existe des techniques différentes. Elles peuvent se regrouper en trois familles distinctes : indépendance, séparation et suffisance. L'objectif des mesures d'indépendance est d'évaluer si les résultats sont indépendants de l'appartenance à un groupe. L'objectif des mesures de séparation est d'évaluer si les résultats sont indépendants de l'appartenance à un groupe, mais conditionnés par les valeurs cibles (ou étiquettes, *labels*) dans les données. L'objectif des mesures de suffisance est d'évaluer si les valeurs cibles sont indépendantes de l'appartenance à un groupe en fonction des résultats. Tous ces types de mesures statistiques sont des moyens valables mais distincts d'interpréter l'équité entre les groupes. Ils présentent tous également des avantages et des limites, et le type de mesure ainsi que la manière d'utiliser les valeurs cibles dans l'évaluation dépendent du contexte. Pour avoir plus des détails sur ces approches, nous conseillons de consulter Castelnovo *et al.* (2022).

Dans le cadre de la prédiction de la réussite académique, quelques travaux se sont concentrés sur l'évaluation de l'équité algorithmique par des approches statistiques. C'est le cas notamment de Gardner *et al.* (2019), Hu et Rangwala (2020), et Lee et Kizilcec (2020), dont les études, menées exclusivement aux Etats-Unis, ont concerné également des modèles de classification binaire pour cette tâche.

Ces travaux ont utilisé différentes mesures d'équité (Caton et Haas, 2020; Verma et Rubin, 2018), comme l'égalité de la précision des prédictions d'un modèle entre les différents groupes (indépendance) ou l'égalité du taux d'erreur de prédiction (séparation) (Baker et Hawn, 2022). Un modèle était alors considéré équitable s'il produisait des performances prédictives similaires entre tous les groupes. Cependant, évaluer l'équité algorithmique à partir des performances prédictives des modèles n'est pas synonyme d'absence de discrimination algorithmique. En effet, un modèle peut produire des erreurs en même quantité pour différents groupes mais elles peuvent être plus sévères pour un groupe que pour un autre car plus éloignées de la réalité et donc avec des implications plus négatives. C'est pourquoi, et en réponse à l'écueil mentionné en Section 1, nous avons développé une nouvelle mesure, la MADD (Verger, Lallé *et al.*, 2023), indépendante de la performance prédictive, que nous utilisons à nouveau ici pour conduire des analyses d'équité algorithmique supplémentaires à celles de Verger, Bouchet *et al.* (2023). Cette mesure sera présentée en Section 3.

De plus, ce sont souvent les mêmes attributs sensibles qui sont étudiés, à savoir le genre et l'origine ethnique (Baker et Hawn, 2022). En effet, Gardner *et al.* (2019) ont comparé les performances prédictives des modèles utilisés par rapport au genre, et Hu et Rangwala (2020) ou Lee et Kizilcec (2020) par rapport au genre et à l'origine ethnique des apprenants. Bien qu'ils soient pertinents à étudier, Baker et Hawn (2022) appellent à considérer une

plus grande diversité d'attributs sensibles pour obtenir davantage d'informations sur leur influence dans l'équité algorithmique. Par conséquent, nos analyses effectuées avec le corpus OULAD (Kuzilek *et al.*, 2017), à la fois dans ce présent article et dans Verger, Bouchet *et al.* (2023), nous permettent d'étudier quatre attributs sensibles : le genre, l'âge, le niveau de pauvreté et le handicap ; les trois derniers ayant été très rarement considérés dans le contexte des analyses d'équité algorithmique en éducation.

2.3. ÉQUITÉ ET DISCRIMINATIONS INTERSECTIONNELLES

Jusqu'à très récemment, toutes les études d'équité algorithmique en éducation proposaient une évaluation attribut sensible par attribut sensible (Baker et Hawn, 2022). Par exemple, le score d'équité du genre était calculé, puis celui de l'origine ethnique, indépendamment. Si cette approche d'évaluation individuelle permet bien de classer les attributs selon leur score obtenu avec n'importe quelle mesure d'équité algorithmique, elle suppose que ces attributs soient indépendants et que les discriminations le soient également. Or, comme mentionné en introduction (Section 1), d'après la théorie de l'intersectionnalité de Crenshaw (1989), les influences conjointes de plusieurs attributs produisent des discriminations uniques et différentes pour certains groupes intersectionnels. Par exemple, les discriminations expérimentées en tant que femmes noires diffèrent des discriminations soit entre hommes et femmes soit entre personnes noires et blanches respectivement (Buolamwini et Gebru, 2018 ; Evans-Winters, 2021). Ainsi, au lieu de calculer un score pour le genre et pour l'origine ethnique séparément, on voudrait un score qui prenne en compte, par exemple, le fait d'être à la fois une femme et d'une certaine origine ethnique.

À notre connaissance, il n'existe qu'une seule étude d'équité algorithmique en éducation qui présente une analyse au niveau de groupes intersectionnels (Zambrano *et al.*, 2024). Dans cette étude, l'équité de deux modèles prédictifs est étudiée pour différents attributs démographiques, incluant l'origine ethnique, le handicap, le genre, la langue maternelle et le niveau de pauvreté, sur une population d'apprenants de plusieurs lycées d'une petite ville du Nord-Est des États-Unis. L'analyse de l'équité est faite en calculant la performance des modèles prédictifs pour tous les groupes possibles à l'intersection entre l'origine ethnique d'un côté, et les autres attributs de l'autre, par exemple les « hispaniques handicapés » ou les « métisses pauvres ». De cette manière, les auteurs identifient les sous-groupes pour lesquels les performances des modèles sont sensiblement plus mauvaises que pour les autres. Leurs résultats ne permettent cependant pas d'identifier de discriminations pour aucun des attributs ni aucune de leurs intersections, peut-être en raison de la taille restreinte de leur base de données (5 000 lycéennes et lycéens provenant d'une même ville). Dans notre étude, nous proposons, comme Zambrano *et al.* (2024), une analyse des discriminations algorithmiques au niveau des attributs seuls (voir Section 6.1) et une autre au niveau des groupes intersectionnels (voir Section 6.2). Nous utiliserons pour cela le corpus OULAD qui est plus large que le corpus utilisé par Zambrano *et al.* (2024) et pour lequel nous avons déjà montré la présence de discriminations algorithmiques pour plusieurs attributs pris séparément. Ce corpus présente donc un potentiel intéressant pour l'étude des groupes intersectionnels, et nos résultats mettent justement en lumière l'utilité de l'approche intersectionnelle sur ce corpus. De plus, nous utiliserons une métrique d'équité algorithmique (la MADD, comme évoqué précédemment en Section 2.2), plutôt que de simplement comparer les performances prédictives des modèles comme dans l'étude de Zambrano *et al.* (2024), afin de pouvoir quantifier les possibles discriminations entre plusieurs groupes intersectionnels.

3. LA MADD, MESURE D'ÉQUITÉ ALGORITHMIQUE

A présent, nous présentons la mesure d'équité algorithmique MADD⁵, utilisée à la fois pour les analyses selon les attributs sensibles individuels en Section 6.1, et pour les analyses selon les groupes intersectionnels réalisées en Section 6.2. En complément des explications présentes dans cette Section 3, nous informons les lectrices et lecteurs que l'article de Verger, Bouchet *et al.* (2023) décrit de manière additionnelle comment exploiter l'aspect visuel de la MADD pour des analyses d'équité, ce dont nous ne nous servons pas ici car nous chercherons à quantifier les discriminations.

3.1. PRÉLIMINAIRES

Considérons un modèle de classification binaire \mathcal{C} pour la prédiction de la réussite à un cours en ligne. Pour pouvoir calculer la MADD, le modèle \mathcal{C} doit fournir pour chaque prédiction, soit une estimation de sa probabilité pour les modèles probabilistes (par exemple : réseaux bayésiens), soit un score de confiance pour les modèles non probabilistes (par exemple : arbres de décision), les deux étant représentés par une valeur comprise entre 0 et 1 inclus.

Par ailleurs, par simplification, nous utiliserons les termes *probabilités prédites* ou *probabilités* pour faire référence à la fois aux estimations de probabilité ou aux scores de confiance. Par exemple, avec un seuil de classification fixé à 0,5, un modèle prédit la réussite (valeur 1) s'il produit une *probabilité* supérieure à 0,5, sinon, il prédit l'échec.

3.2. EXPLICATIONS

Supposons que le modèle \mathcal{C} prédise des probabilités de réussite comme illustrées en Figures 1a et 1b, pour deux groupes d'apprenants distincts G_1 et G_2 . Ces deux groupes peuvent être des groupes distincts d'apprenants, c'est-à-dire qu'un apprenant ne peut pas appartenir aux deux groupes simultanément. Ces groupes peuvent être issus d'un même attribut (par exemple les hommes *vs.* les femmes pour l'attribut de genre) comme des groupes intersectionnels distincts (pour les attributs de genre et de handicap, les femmes avec un handicap forment un groupe distinct des hommes eux-mêmes avec un handicap).

Les histogrammes représentent alors la distribution des probabilités de réussite produites par le modèle \mathcal{C} , pour chacun de ces groupes. En Section 3.3, nous appellerons une telle distribution "vecteur de densité" des probabilités. Chaque barre verticale décrit la proportion d'apprenants ayant reçus la même probabilité de réussite. A titre d'exemple, sur ces histogrammes nous pouvons constater que les probabilités de G_1 sont surtout situées entre 0 et 0,5, alors que celles de G_2 sont plus élevées, entre 0,5 et 0,7 environ. Le modèle a donc tendance à donner de meilleures probabilités de réussite à G_2 qu'à G_1 .

Ainsi, pour quantifier la différence de comportement du modèle \mathcal{C} entre les deux groupes G_1 et G_2 , la MADD a été conçue (Verger, Lallé *et al.*, 2023) spécifiquement pour mesurer les différences entre les distributions que les histogrammes représentent. Nous en proposons une illustration en Figure 1c. Pour cela, nous avons illustré par des lignes courbes, les estimations continues des distributions représentées par les histogrammes à l'aide d'une méthode d'estimation de densité par noyau (ou *kernel density estimation*). La différence entre ces deux courbes, illustrée par une zone rouge dans laquelle le modèle ne produit pas les mêmes probabilités de réussite pour les deux groupes, est donc ce que la MADD cherche à quantifier. Nous présentons sa définition dans la section suivante. Il est important de souligner que la définition de la MADD ne s'appuie pas sur les estimations continues utilisées pour l'illustration mais bien sur les probabilités effectivement prédites par le modèle \mathcal{C} .

5. Traduisible en "Distance Absolue entre les Densités du Modèle".

3.3. DÉFINITION

Soient les vecteurs de densité $D^{G1} = [d_0^{G1}, d_1^{G1}, \dots, d_m^{G1}]$ et $D^{G2} = [d_0^{G2}, d_1^{G2}, \dots, d_m^{G2}]$, associés aux groupes G1 et G2 respectivement et illustrés par les Figures 1a et 1b. Les valeurs d_k avec $0 \leq k \leq m$ représentent les proportions de chaque probabilité discrète obtenue, c'est-à-dire la proportion de chaque barre verticale des Figures 1a et 1b, et où m correspond au nombre de probabilités discrètes considérées (Verger, Lallé *et al.*, 2023). Comme chaque vecteur représente la proportion totale des probabilités reçues dans chaque groupe, la somme de leurs éléments vaut toujours 1 respectivement. Ainsi, la MADD est définie comme suit :

$$\text{MADD}(D^{G1}, D^{G2}) = \sum_{k=0}^m |d_k^{G1} - d_k^{G2}| \quad (1)$$

La MADD est par conséquent 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 apprenants de G1 (i.e., proportion maximale pour une seule valeur de probabilité donnée) et une probabilité de p_j (avec $p_j \neq p_i$) à tous les apprenants de G2. Ainsi, pour n'importe quelles probabilités indexées par i et j :

$$\text{MADD}(D^{G1}, D^{G2}) = |d_i^{G1}| + |d_j^{G2}| = (1 + 1) = 2 \quad (2)$$

Par ailleurs, il existe un nombre optimal m de probabilités discrètes (ou de barres dans les histogrammes) à considérer pour le calcul de la MADD. Avec ce nombre optimal, la MADD présente des garanties théoriques de sa précision pour la mesure de la différence entre les deux distributions de chaque groupe, ce qui est montré dans Verger *et al.* (2024). Les expériences présentées dans les sections suivantes ont été réalisées avec ce nombre optimal. Cependant, la détermination de celui-ci n'est pas l'objet de cet article et nous renvoyons les lectrices et lecteurs au travail de Verger *et al.* (2024) quant à son choix et son usage.

4. ANALYSE DES GROUPES INTERSECTIONNELS

Dans cette section, nous revenons d'abord plus en détails sur des exemples de groupes intersectionnels en Section 4.1. Puis, nous présentons dans la Section 4.2 les différentes

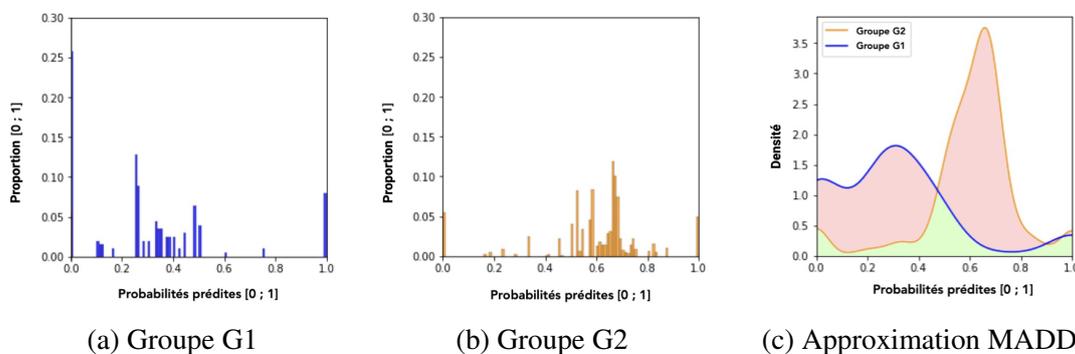


Figure 1 : Histogrammes des probabilités prédites pour deux groupes distincts (a, b) et représentation visuelle de la MADD (c)

approches possibles pour déterminer des groupes G1 et G2 intersectionnels (voir Equation 1) à analyser. Ces approches sont applicables avec toute mesure d'équité de groupe telle que la MADD et avec celles mentionnées en Section 2. Enfin, nous discutons du choix de ces approches en Section 4.3.

4.1. EXEMPLES DE GROUPES INTERSECTIONNELS

Comme mentionné auparavant, un groupe intersectionnel est un groupe à l'intersection de plusieurs attributs. De tels groupes peuvent être représentés comme en Figure 2 (Yang *et al.*, 2020), où ils sont identifiés par les lettres A, B, C, D, E, F, G, H. À titre d'exemple, le groupe A inclut les femmes plus âgées avec un handicap. L'ensemble des femmes est alors l'union des groupes intersectionnels $A \cup B \cup E \cup F$, l'ensemble des personnes plus âgées $A \cup C \cup E \cup G$ et l'ensemble des personnes avec un handicap $A \cup B \cup C \cup D$.

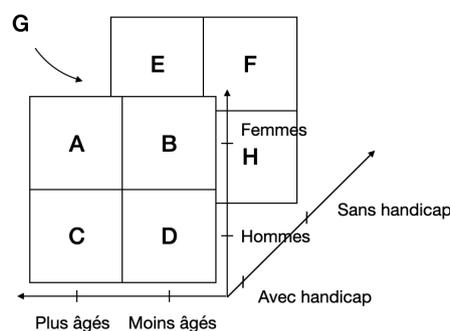


Figure 2 : Représentation des groupes intersectionnels avec trois attributs binaires (âge, genre et handicap)

4.2. CHOIX DES GROUPES INTERSECTIONNELS DE COMPARAISON

À l'inverse de l'analyse par attribut individuel, conduite usuellement dans la littérature (voir Section 2) et plus loin dans l'article (Section 6.1), consistant à comparer les groupes qui composent un attribut donné (par exemple, le groupe avec handicap et le groupe sans handicap pour l'attribut handicap), nous dénombrons deux manières de comparer les groupes intersectionnels, présentées ci-dessous.

4.2.1. Un groupe intersectionnel vs. le reste

Premièrement, il est possible de comparer un groupe intersectionnel (G1) avec le reste des apprenants dans les données (G2). Par exemple sur la Figure 2, le groupe A (femme plus âgées avec handicap) peut être comparé à l'union de tous les autres groupes (B à H). Ceci permet d'évaluer si un groupe donné est plus ou moins discriminé par rapport au reste des individus (hommes, femmes moins âgées sans handicap, femmes moins âgées avec handicap). Nous proposons une telle analyse en Section 6.2.4, conduisant à une visualisation sous forme d'histogrammes.

4.2.2. Un groupe intersectionnel vs. un autre groupe intersectionnel

Deuxièmement, il est possible de comparer un groupe intersectionnel (G1) vis-à-vis d'un autre groupe intersectionnel (G2). Par exemple, sur la Figure 2, nous pouvons comparer les groupes A et B entre eux. Cela permet d'évaluer si un groupe est plus ou moins discriminé

par rapport à un autre groupe spécifique. Nous avons également réalisé de telles analyses dans les Sections 6.2.2 et 6.2.3, conduisant à une visualisation sous forme de matrice.

4.3. DISCUSSION

Nous pouvons constater qu’avec la première approche, il est aisé de comparer tous les groupes intersectionnels un par un avec le reste des individus. De cette manière, nous pourrions ordonner ces résultats pour trouver les groupes intersectionnels conduisant aux résultats de MADD les plus sévères, et inversement. Cela permet donc d’identifier des groupes intersectionnels particuliers.

Par contre, avec la deuxième approche, il n’est pas forcément évident de savoir entre quels groupes intersectionnels effectuer la comparaison. Nous pouvons comparer chaque groupe intersectionnel avec tous les autres un par un (voir Figure 6). Mais certaines comparaisons, entre des groupes intersectionnels spécifiques, permettent de connaître l’impact discriminant d’un attribut sur un autre. En effet, en Figure 3, nous illustrons les deux types de comparaisons possibles entre deux attributs, en l’occurrence les attributs de genre et d’âge. Nous pouvons observer que dans la configuration de gauche en Figure 3, nous comparons l’équité algorithmique entre les groupes A-C, A-D, B-C, et B-D. De cette manière, les femmes sont toujours comparées aux hommes, en particulier les femmes avec une distinction d’âge et les hommes avec une distinction d’âge. Dans cette configuration, l’impact discriminant de l’âge sur le genre peut donc être mis en évidence. À l’inverse, dans la configuration de droite en Figure 3, les personnes plus âgées sont comparées avec celles moins âgées, en ajoutant une distinction de genre. Cette fois, c’est l’effet de l’appartenance d’un groupe du genre qui est évalué sur celui de l’âge. Ce sont donc les comparaisons A-B, A-D, C-B et C-D qui sont en jeu. Ainsi, seules les comparaisons A-D et B-C sont communes entre ces deux types de comparaisons et, par conséquent, selon que l’on souhaite évaluer dans un sens ou dans l’autre l’impact d’un attribut sur un autre, l’une ou l’autre des configurations sera préférée. Nous allons mettre en pratique cette approche en Section 6.2.3.

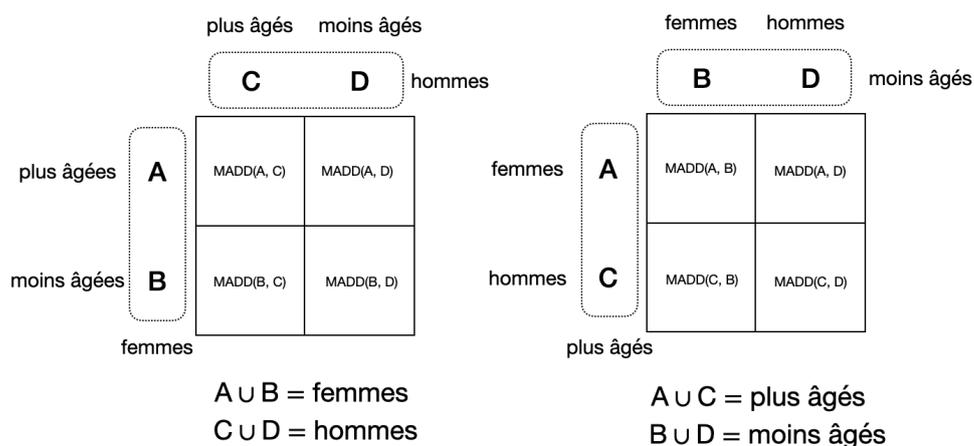


Figure 3 : Deux types de comparaisons pour évaluer l’impact discriminant d’un attribut sur un autre, avec l’exemple de deux attributs binaires (âge, genre)

5. EXPÉRIENCES

Dans les sections suivantes, nous allons présenter les données utilisées pour nos expériences (Section 5.1), les attributs sensibles considérés (Section 5.2) et les modèles étudiés pour les analyses (Section 5.3).

5.1. CORPUS DE DONNÉES OULAD

Nous conduisons nos analyses d'équité algorithmique à partir du jeu de données OULAD (*Open University Learning Analytics Dataset* - Kuzilek *et al.* (2017)). En effet, il s'agit d'un corpus anonymisé largement utilisé en éducation (Alhakbani et Alnassar, 2022), notamment pour la prédiction de la réussite à des cours en ligne. De plus, 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 (Hellas *et al.*, 2018). Aussi, il contient des données de différents cours avec des profils d'apprenants variés, ce qui nous permet de répliquer nos expériences dans plusieurs contextes avec des populations différentes (autre appel lancé par Hellas *et al.* (2018)). Enfin, 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 suivis sans prérequis de manière indépendante ou dans le cadre d'un cursus universitaire. Les apprenants étaient inscrits 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). Ces informations sont regroupées dans le Tableau 1.

Tableau 1 : Informations sur le jeu de données OULAD

Cours	Domaine	Nombre d'apprenants
AAA	Sciences sociales	748
BBB	Sciences sociales	7 909
CCC	STIM	4 434
DDD	STIM	6 272
EEE	STIM	2 934
FFF	STIM	7 762
GGG	Sciences sociales	2 534

Le corpus contient des données démographiques et des données d'activité dans l'espace numérique de travail (ENT), avec initialement 32 593 échantillons (paire apprenant - cours). Nous avons utilisé les attributs présentés dans le Tableau 2 ainsi que la variable cible binaire « réussite/échec ». 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.

5.2. ATTRIBUTS SENSIBLES ET SÉLECTION DES COURS

Nous ciblons dans nos expériences l'étude du caractère sensible des 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, à condition

Tableau 2 : Attributs utilisés du jeu de données OULAD

Attribut	Type	Description
genre	binaire	genre de l'apprenant
age	ordinal	intervalle de l'âge de l'apprenant
handicap	binaire	indique si l'apprenant a déclaré un handicap
dernier_diplome	ordinal	dernier diplôme de l'apprenant à l'entrée du cours
pauvrete	ordinal	niveau de pauvreté du lieu d'habitation de l'apprenant
nb_tentatives	numérique	nombre de tentatives précédentes au cours
credits	numérique	nombre de crédits pour le cours étudié
nb_total_click	numérique	nombre total d'interactions de l'apprenant avec le cours

qu'ils soient binaires ou catégoriels pour la définition des groupes intersectionnels. Cependant, ici nous allons plutôt chercher à confirmer ou infirmer le caractère sensible de ces quatre attributs retenus. Par ailleurs, pour distinguer deux groupes G1 et G2 pour les attributs `pauvrete` et `age` dans leur analyse de manière individuelle (voir Section 6.1), nous utilisons un seuil de 50% de l'indice de pauvreté britannique (Kuzilek *et al.*, 2017) et nous distinguons, parmi les trois tranches d'âge disponibles dans les données ([0-35] : 13 815 individus, ou 69% ; [36-55] : 6 011 individus, ou 30% et [55+] : 138 individus, ou 18%), le groupe majoritaire ([0-35]) du groupe minoritaire (regroupement de [36-55] et [55+]).

Quant aux cours étudiés, parmi les sept disponibles, nous avons sélectionné dans le corpus le cours de Sciences sociales identifié "BBB" et le cours de STIM identifié "FFF". En effet, d'après une analyse des corrélations des attributs, notamment en Figure 4, ces deux cours ont présenté les plus fortes corrélations avec l'attribut `genre`, ce qui suggère une importance de celui-ci 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 avec d'une part, une large majorité de femmes (91,2 %) dans le cours "BBB", et d'autre part, une majorité d'hommes dans le cours "FFF" (88,4 %). Par ailleurs, ces deux cours présentent l'avantage d'être dans deux domaines différents et d'avoir les effectifs d'apprenants les plus élevés (voir Tableau 1). Ainsi, ces deux cours sont de très bons candidats pour notre analyse des discriminations algorithmiques par rapport aux attendus de biais de genre notamment.

5.3. MODÈLES PRÉDICTIONNELS DE LA RÉUSSITE ÉTUDIÉE

Dans un souci de généralisation, nous avons sélectionné plusieurs types de modèles de classification pour nos analyses, respectivement à base de régressions, 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 (Alhakbani et Alnassar, 2022 ; Korkmaz et Correia, 2019), 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 de complexité supérieure 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.

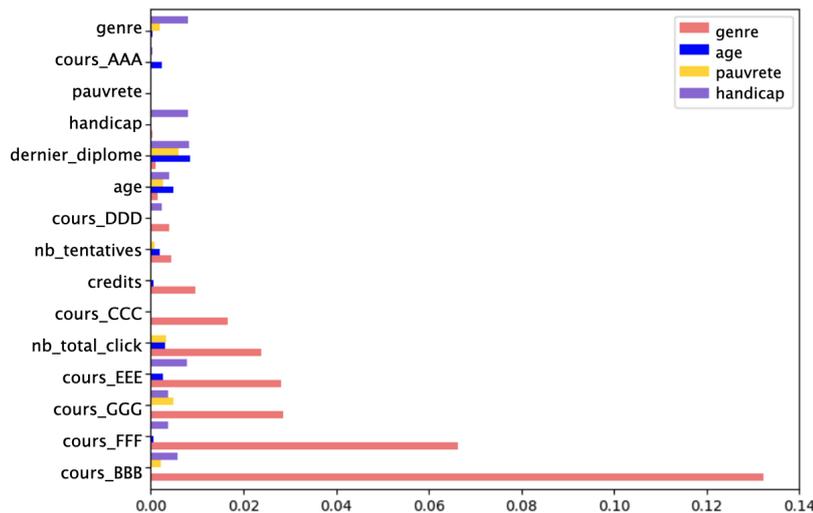


Figure 4 : Scores d'informations mutuelles (Kullback, 1959) entre les attributs de gauche avec ceux en légende

Concernant l'entraînement des modèles retenus, nous avons utilisé 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 de données à l'aide d'une sélection aléatoire stratifiée. 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 que, contrairement aux études qui se concentrent sur l'apprentissage automatique, l'objectif ici n'est pas d'obtenir les meilleures performances prédictives mais d'illustrer l'intérêt de nos méthodes sur divers modèles. Ce point serait en revanche important dans le cas du déploiement réel d'un modèle prédictif. Enfin, les analyses et les évaluations avec le calcul de la MADD ont été réalisées sur le jeu de test. Ces évaluations peuvent être reproduites grâce aux données et au code documenté, disponibles à l'adresse indiquée à la note de pied de page n°3.

6. RÉSULTATS

Dans cette section, nous présentons les analyses d'équité algorithmique, d'une part, selon les attributs sensibles individuellement en Section 6.1, et d'autre part, selon les attributs sensibles simultanément grâce aux groupes intersectionnels en Section 6.2. Les premières ont été en partie reprises des travaux précédents (Verger, Bouchet *et al.*, 2023) afin d'être comparées aux deuxièmes.

6.1. ANALYSES D'ÉQUITÉ PAR ATTRIBUTS SENSIBLES INDIVIDUELS

Les résultats de MADD pour chaque modèle vis-à-vis de chaque attribut sensible sont présentés dans les Tableaux 3 et 4 pour le cours de Sciences sociales et de STIM respectivement. Ces tableaux se lisent comme suit : 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 quant à elles indiquées en **rouge**. Par ailleurs, les lectrices et

lecteurs peuvent consulter Verger, Bouchet *et al.* (2023) pour l'exploitation visuelle des résultats de la MADD pour obtenir davantage de détails.

6.1.1. Cours de Sciences sociales (“BBB”)

Dans le Tableau 3, nous voyons par leur couleur rouge que les résultats de `pauvrete` sont les plus élevés pour trois modèles sur quatre (LR, KN et DT). Ces modèles produisent donc des prédictions de réussite qui diffèrent le plus en fonction de l'appartenance à l'un ou l'autre des groupes de `pauvrete`. L'âge obtient des résultats similaires à `pauvrete`. Ce sont les deux attributs qui ont des moyennes de MADD les plus hautes (1,23 et 1,18).

En revanche, le `genre`, attendu comme l'attribut potentiellement le plus discriminant (voir Section 5.2), s'avère obtenir des valeurs inférieures à celles de `pauvrete` ou de `age`. Enfin, c'est le `handicap`, avec ces trois astérisques sur quatre, qui est l'attribut vis-à-vis duquel les trois modèles LR, KN et DT discriminent le moins. Il obtient d'ailleurs la moyenne la plus basse (0,82).

Tableau 3 : Résultats de la MADD pour le cours “BBB”

	Modèle	Attributs sensibles				Moyenne
		genre	age	handicap	pauvrete	
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	

6.1.2. Cours de STIM (“FFF”)

Pour le cours de STIM, le Tableau 4 montre, qu'en revanche, `pauvrete` est l'attribut pour lequel les modèles discriminent le moins, avec la moyenne la plus basse (0,77). 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'âge est également élevée.

Tableau 4 : Résultats de la MADD pour le cours “FFF”

	Modèle	Attributs sensibles				Moyenne
		genre	age	handicap	pauvrete	
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	

6.1.3. Bilan

Cette approche d'analyse d'équité algorithmique par attributs sensibles individuels nous a permis de déterminer leur pouvoir discriminant, soit en moyenne pour tous les modèles, soit pour chaque modèle en particulier. Cette approche a permis de mettre en exergue que `pauvrete` et `genre` étaient des attributs sensibles cruciaux dans le cours de Sciences

sociales ou de STIM, mais également que `age` avait toujours la moyenne de MADD la plus élevée dans les deux cours. Pour répondre à nos questions de recherche introduites en Section 1, nous allons maintenant étudier quels seraient les résultats d'équité si nous considérions plusieurs attributs sensibles ensemble.

6.2. ANALYSES D'ÉQUITÉ PAR GROUPES INTERSECTIONNELS

Nous nous intéressons ici aux discriminations intersectionnelles présentes dans les résultats, c'est-à-dire aux discriminations qui viennent de l'influence de plusieurs attributs sensibles simultanément. Pour rappel, un groupe intersectionnel est un groupe à l'intersection de plusieurs attributs. Ces analyses nous permettront de comparer l'influence d'attributs sensibles pris seuls dans la section précédente, avec l'influence d'attributs sensibles pris ensemble.

6.2.1. Cadre de ces analyses

Les analyses étant plus nombreuses du fait du nombre important de groupes intersectionnels possibles à partir de quatre attributs (i.e., `genre`, `age`, `handicap`, `pauvrete`), nous allons restreindre cette partie aux résultats du modèle DT (arbre de décision) et au cours "BBB". En effet, concernant le modèle DT, il est un bon candidat pour l'analyse plus approfondie des discriminations algorithmiques du fait de sa bonne précision et de ses bons résultats de MADD : il a obtenu la précision la plus élevée (93%) et il a obtenu les meilleures moyennes de MADD dans les deux cours de Sciences sociales et de STIM (voir les moyennes à droite dans les Tableaux 3 et 4). Par conséquent, il aurait été un modèle de choix dans un scénario réel, ce qui présente un intérêt pour les analyses.

De plus, parmi les résultats du modèle DT, nous nous concentrons sur le cours "BBB" car il contient les deux attributs ayant reçu la MADD la plus élevée parmi les deux cours, à savoir `age` (0,84) et `pauvrete` (0,85). Ainsi, par rapport à une valeur de référence de 0,85 qui est la valeur maximale obtenue par le DT, nous pourrions examiner si un groupe formé à partir, d'une part, de l'attribut `pauvrete` ou `age` et, d'autre part, d'un autre attribut, présente une discrimination encore plus élevée que `pauvrete` ou `age` seul, ou bien si cet autre attribut aura tendance à faire diminuer l'influence du premier.

Évaluer l'impact de `pauvrete` ou `age` sur un autre attribut sera l'objet des Sections 6.2.2 et 6.2.3. Puis, nous ferons une analyse de tous les groupes intersectionnels possibles en Section 6.2.4, pour une recherche exhaustive des discriminations algorithmiques.

6.2.2. L'influence de l'attribut `pauvrete` sur les autres attributs sensibles

Tout d'abord, nous examinons l'influence de `pauvrete` sur les autres attributs. En effet, comme il est individuellement celui qui a obtenu la valeur de MADD la plus élevée, et qu'il est donc l'attribut le plus discriminé ici, il est intéressant d'étudier si, évalué conjointement avec un autre attribut, il forme des groupes intersectionnels davantage discriminés. Les résultats sont présentés dans les Tableaux 5, 6 et 7.

Sur l'ensemble des trois tableaux, aucune valeur n'excède la valeur de référence 0,85, ce qui signifie qu'appartenir à un groupe d'un autre attribut n'aggrave pas les discriminations algorithmiques déjà observées pour l'attribut `pauvrete`, voire les réduit significativement puisque très proches de 0 : 0,11, 0,09, 0,06 ou encore 0,03. De plus, nous constatons que les résultats sur la diagonale (moins pauvre, moins pauvre)-(plus pauvre, plus pauvre) sont bien plus faibles que sur l'autre diagonale (plus pauvre, moins pauvre)-(moins pauvre, plus pauvre), ce qui est illustré avec des pointillés rouges ou oranges pour l'exemple dans le Tableau 5. Cela confirme le poids plus important, dans les discriminations algorithmiques, de

l'attribut `pauvrete` car la comparaison entre les groupes opposés de `pauvrete` cause une augmentation significative de la MADD alors même qu'elle est très faible pour les groupes opposés de `genre`, `handicap`, et `age`. Pour résumer, l'attribut `pauvrete` ne présente pas d'intersectionnalité avec les autres attributs, c'est-à-dire n'engendre pas de discriminations plus importantes par rapport à sa considération seule.

Tableau 5 : Résultats de la MADD pour les groupes intersectionnels de `genre` et `pauvrete`. En pointillés rouges est représentée la diagonale (moins pauvre, moins pauvre)-(plus pauvre, plus pauvre) et en pointillés orange la diagonale (plus pauvre, moins pauvre)-(moins pauvre, plus pauvre).

		Homme	
		Moins pauvre	Plus pauvre
Femme	Moins pauvre	0,11	0,46
	Plus pauvre	0,71	0,19

Tableau 6 : Résultats de la MADD pour les groupes intersectionnels de `handicap` et `pauvrete`

		Handicap	
		Moins pauvre	Plus pauvre
Non handicap	Moins pauvre	0,09	0,29
	Plus pauvre	0,72	0,37

Tableau 7 : Résultats de la MADD pour les groupes intersectionnels de `age` et `pauvrete`

		Moins âgé	
		Moins pauvre	Plus pauvre
Plus âgé	Moins pauvre	0,03	0,63
	Plus pauvre	0,69	0,06

6.2.3. L'influence de l'attribut `age` sur les autres attributs sensibles

L'autre attribut ayant obtenu la deuxième valeur la plus élevée de MADD, `age`, ne produit quant à lui pas les mêmes résultats. En effet, les Tableaux 8, 9 et 10 montrent que la MADD dépasse la valeur de référence (0,85) avec les trois autres attributs. Les groupes concernés sont le groupe des femmes plus âgées contre les hommes moins âgés (0,93), le groupe des personnes plus âgées déclarées sans handicap contre les personnes moins âgées déclarées avec handicap (0,91) et le groupe des personnes moins pauvres plus âgées contre les personnes plus pauvres moins âgées (1,15). Cela montre, par les mêmes croisements au niveau des diagonales, que pour l'attribut `pauvrete`, l'attribut `age` a un poids plus important que `genre`, `handicap` ou même `pauvrete` dans les discriminations algorithmiques intersectionnelles. Autrement dit, dans le cas analysé dans le Tableau 8, il y a une discrimination plus importante entre les femmes plus âgées et les hommes plus âgés que les personnes (hommes et femmes) âgées vs. les personnes moins âgées. Pour résumer, l'attribut `age` présente une intersectionnalité avec les autres attributs, c'est-à-dire qu'il engendre des discriminations plus importantes par rapport à sa considération seule.

Tableau 8 : Résultats de la MADD pour les groupes intersectionnels de genre et age

		Homme	
		Moins âgé	Plus âgé
Femme	Moins âgé	0,06	0,66
	Plus âgé	0,93	0,32

Tableau 9 : Résultats de la MADD pour les groupes intersectionnels de handicap et age

		Handicap	
		Moins âgé	Plus âgé
Non handicap	Moins âgé	0,05	0,65
	Plus âgé	0,91	0,33

Tableau 10 : Résultats de la MADD pour les groupes intersectionnels de pauvreté et age

		Plus pauvre	
		Moins âgé	Plus âgé
Moins pauvre	Moins âgé	0,12	0,69
	Plus âgé	1,15	0,36

6.2.4. Tous les groupes intersectionnels

Enfin, nous examinons les discriminations algorithmiques avec tous les groupes intersectionnels possibles, à savoir seize (2^4) puisque nous avons quatre attributs sensibles binaires à disposition. Ces groupes intersectionnels sont recensés dans le Tableau 11, avec les alias que nous leur avons attribué pour l'analyse des figures à venir, ainsi que leur effectif. Nous pouvons soit calculer l'équité de chaque groupe contre le reste des apprenants (cf. Figure 5), soit calculer l'équité groupe contre groupe (cf. Figure 6), comme introduit en Section 4.

GROUPES INTERSECTIONNELS VS. LE RESTE. Dans la Figure 5, chaque barre verticale représente la MADD d'un groupe intersectionnel contre le reste des individus. La ligne pointillée rouge, indiquant la valeur de référence (0.85), montre qu'un groupe en particulier dépasse cette valeur : les hommes plus pauvres avec un handicap et plus âgés (alias A, combinaison 1-1-1-1). Du point de vue des effectifs, comme reporté dans le Tableau 11, le groupe A est très minoritaire puisque constitué de seulement 3 apprenants, mais obtient tout de même une MADD au moins deux fois supérieure à d'autres groupes minoritaires comparables comme les groupes B, E et F avec respectivement 10, 1 et 5 apprenants. La question des effectifs sera soulevée en discussion dans la Section 7. Par ailleurs, nous pouvons également observer quatre barres aux environs de 0,60 qui se distinguent des autres, les groupes B, E, I et M, dont le point commun est d'être des groupes dans lesquels les personnes ont déclaré un handicap. Ceci n'aurait pas pu être relevé sans cette analyse des discriminations intersectionnelles.

GROUPES INTERSECTIONNELS VS. GROUPES INTERSECTIONNELS. Dans la Figure 6, chaque case de la matrice contient un cercle dont le diamètre représente la valeur de MADD de la comparaison d'un groupe intersectionnel en ordonnées contre un groupe intersectionnel en abscisses. La lecture de la figure se fait donc de la gauche vers le bas. La diagonale étant les comparaisons de chaque groupe avec lui-même, les cases correspondantes contiennent bien une MADD de valeur nulle. De plus, en nommant à nouveau deux groupes

Tableau 11 : Alias de tous les groupes intersectionnels. Pour genre, les hommes sont désignés par la valeur 1 et les femmes par la valeur 0. Pour pauvreté, les plus pauvres sont désignés par la valeur 1 et les moins pauvres par la valeur 0. Pour handicap, ceux déclarés avec un handicap sont désignés par la valeur 1 et ceux qui n'ont déclaré aucun handicap par la valeur 0. Enfin, pour âge, les personnes plus âgées sont désignées par la valeur 1 et les personnes moins âgées par la valeur 0.

Alias	genre	pauvrete	handicap	age	Effectif
A	1	1	1	1	3
B	1	1	1	0	10
C	1	1	0	1	44
D	1	1	0	0	44
E	1	0	1	1	1
F	1	0	1	0	5
G	1	0	0	1	35
H	1	0	0	0	44
I	0	1	1	1	31
J	0	1	1	0	48
K	0	1	0	1	234
L	0	1	0	0	514
M	0	0	1	1	17
N	0	0	1	0	27
O	0	0	0	1	210
P	0	0	0	0	323

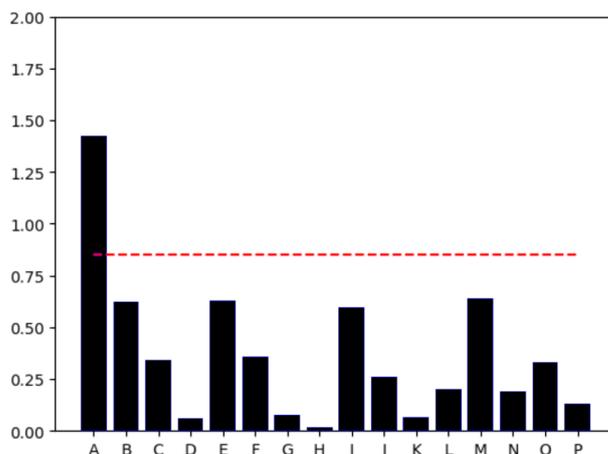


Figure 5 : Résultats de la MADD pour tous les groupes intersectionnels vis-à-vis du reste des apprenants

intersectionnels par G1 et G2, il est possible de comparer G1 avec G2 ou bien G2 avec G1. C'est pourquoi la matrice est symétrique. Nous avons également ajouté des couleurs pour préciser le sens des comparaisons : en lisant toujours de gauche à en bas et en prenant, par exemple, le groupe C en ordonnées, C présente un cercle vert avec A, signifiant qu'il est favorisé par le modèle par rapport à A, tandis qu'il présente un cercle rouge avec E, signifiant qu'il est plus discriminé par le modèle par rapport à E.

Dans cette matrice, nous pouvons constater deux choses. Premièrement, il y a 2 cases

avec une MADD maximale de 2 (i.e., cercle de diamètre maximal). Ce sont les comparaisons des groupes :

- (A) les hommes plus pauvres avec un handicap et plus âgés,
- (E) les hommes moins pauvres avec un handicap et plus âgés (pauvreté diffère),

et de :

- (A) les hommes plus pauvres avec un handicap et plus âgés,
- (M) les femmes moins pauvres avec un handicap et plus âgées (genre et pauvreté différent).

Dans ces deux comparaisons A-E et A-M avec des effectifs respectifs de 3-1 et 3-17, nous tombons dans le cas où la MADD est maximale car le modèle ne produit aucune probabilité commune entre les deux groupes puisqu’il y a de toute façon très peu de probabilités concernées. A nouveau, la question des effectifs sera abordée dans la Section 7.

Deuxièmement, nous constatons que la ligne du groupe A semble celle ayant les valeurs globales les plus élevées, plus que le groupe E constitué pourtant d’un seul apprenant. Ce groupe A semble donc être le groupe envers lequel il est nécessaire d’être le plus vigilant dans le cours de Sciences sociales “BBB”, du fait de son effectif très réduit et des probabilités très faibles qu’il reçoit (tous ses cercles sont rouges et de diamètre élevé). D’autres lignes avec une très grande majorité de cercles rouges peuvent également attirer notre attention comme les groupes B, C, I et J.

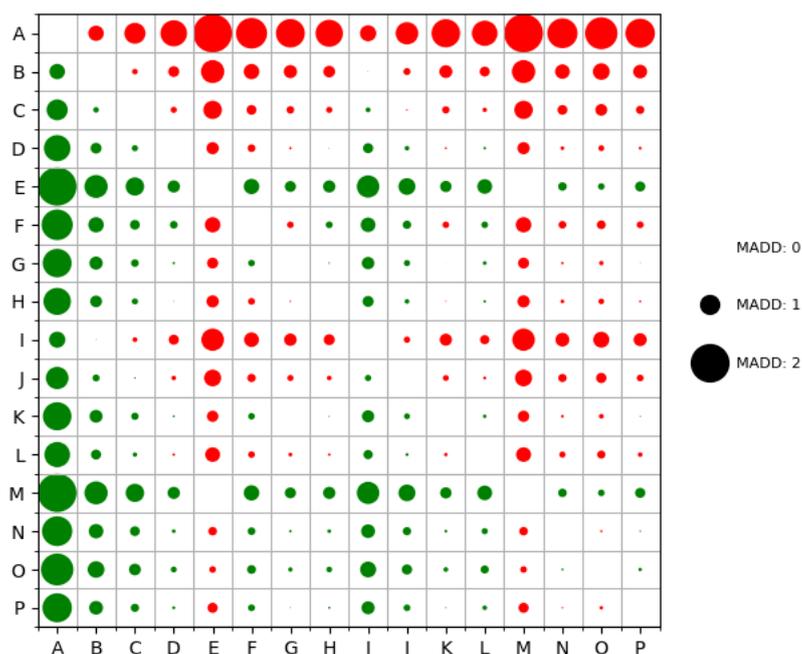


Figure 6 : Résultats de la MADD des groupes intersectionnels entre eux

6.2.5. Bilan

Grâce aux analyses d’équité algorithmique par groupes intersectionnels, nous avons pu comprendre de manière plus fine l’influence de chaque attribut sensible par leurs interactions avec les autres. Cela nous a également permis de découvrir des discriminations invisibles avec une analyse par attributs sensibles individuels, comme celles pour les groupes A, E et M, et en particulier celui concernant les hommes plus pauvres avec un handicap et plus âgés (groupe A). Une question peut donc émerger : comment ne pas passer à côté de toutes les

discriminations possibles ? Autrement dit, est-il possible de détecter automatiquement les bons groupes intersectionnels à analyser ? Ce point sera abordé ci-après.

7. DISCUSSION

Dans cette section, nous allons d’abord soulever les limites des données et modèles considérés (Section 7.1), puis les points clés des résultats obtenus avec les attributs sensibles individuels (Section 7.2) et avec les groupes intersectionnels (Section 7.3). Ensuite, nous discuterons du choix des attributs sensibles et des précautions à prendre vis-à-vis de leurs effectifs (Section 7.4). Enfin, nous proposerons quelques recommandations pratiques (Section 7.5).

7.1. DONNÉES ET MODÈLES ÉTUDIÉS

Le jeu de données OULAD contient peu d’attributs et presque tous catégoriels. Par conséquent, il présente une diversité et une variance très faibles dans les valeurs des attributs. Pour cette raison, tout modèle appliqué sur ces données aura des difficultés à distinguer avec précision les étudiants qui réussissent et ceux qui échouent aux cours. OULAD, qui est l’un des rares jeux de données ouverts comportant plusieurs attributs sensibles, présente donc un désavantage significatif sur la qualité limitée des données et la représentation réduite des individus, ce qui aurait pu permettre une évaluation plus robuste de l’équité algorithmique.

Aussi, dans nos expériences, nous nous sommes délibérément concentrés sur les résultats de la MADD pour les analyses d’équité. Cependant, pour des applications réelles, il est important de prêter attention à la fois à la performance et à l’équité afin de sélectionner des modèles pertinents pour leur usage. Par exemple, le modèle NB dans nos expériences pourrait être considéré comme équitable au regard de ses résultats de MADD, mais il présentait une précision très faible. En particulier, il avait des difficultés à prédire correctement la réussite ou l’échec des individus par rapport à tous les attributs, pas que sensibles, ce qui rendrait ce modèle inutile pour des cas réels, mais néanmoins intéressant ici pour notre analyse exploratoire. Notamment, nous n’avons effectué aucun ajustement des modèles (ou *fine tuning*) pour illustrer un cas général avec différents modèles. Il convient à l’inverse, d’étudier attentivement cette phase du processus. Nous recommandons donc d’utiliser la MADD pour des modèles qui montrent une performance prédictive satisfaisante concernant le problème en jeu, afin d’acquérir une compréhension plus fine de la façon dont ils se comportent et vis-à-vis de qui, et d’affiner leur sélection.

De plus, utiliser un modèle prédictif qui serait biaisé, pour une application réelle (et potentiellement à grande échelle), pourrait perpétuer ou produire des discriminations envers certains individus. L’enjeu est donc de systématiquement évaluer la performance et l’équité des modèles en profondeur pour être informé de ses potentiels biais, mais aussi pour anticiper les conséquences négatives de leur utilisation.

7.2. RÉSULTATS AVEC ATTRIBUTS SENSIBLES INDIVIDUELS

Les résultats au niveau des attributs sensibles ont permis de montrer qu’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. En effet, 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 à l’origine des discriminations algorithmiques les plus importantes. En revanche, cela n’est pas le cas pour le cours de STIM qui, lui, présente des résultats de MADD plutôt alignés avec le biais de genre dans les données et des plus faibles discriminations algorithmiques selon la pauvreté.

7.3. RÉSULTATS AVEC GROUPES INTERSECTIONNELS

Les résultats au niveau des groupes intersectionnels ont permis de mettre en lumière des discriminations qu'on ne pouvait voir avec l'analyse des attributs individuels seuls et de comprendre de manière plus fine l'influence de chaque attribut sur les discriminations grâce à leurs interactions avec les autres attributs. En effet, ces analyses ont permis de découvrir des groupes particuliers, dont celui des hommes plus pauvres avec un handicap et plus âgés, et de montrer que l'attribut d'âge engendrait une intersectionnalité avec les autres attributs, à l'inverse de l'attribut pauvreté. L'analyse par groupes intersectionnels est donc importante pour l'équité algorithmique, et a ainsi permis de répondre à la QR2, sur la découverte de discriminations algorithmiques supplémentaires en considérant les individus à l'intersection de plusieurs attributs sensibles.

7.4. CHOIX ET EFFECTIFS DES GROUPES INTERSECTIONNELS

Dans la Section 4, nous avons présenté et discuté de deux approches pour étudier l'équité algorithmique vis-à-vis de plusieurs attributs sensibles. Ainsi, nous avons pu apporter une réponse à la QR1 qui était de savoir comment évaluer l'influence de plusieurs attributs sensibles simultanément. Nous pouvons en effet comparer un groupe intersectionnel avec le reste des individus, ou bien un groupe avec un autre groupe, dont certaines comparaisons de groupes peuvent faire ressortir l'influence d'un attribut sensible sur un autre.

Par ailleurs, il est à souligner que les groupes intersectionnels peuvent être formés à partir d'attributs binaires, comme dans cet article, ou catégoriels de manière générale, et que, pour des attributs numériques, une question de seuil se pose (comment faire des catégories pertinentes à partir de valeurs continues?). Différents attributs conduisent à différents groupes intersectionnels. Il y a donc un choix à faire, qui reste empirique, des attributs à considérer, et donc des groupes intersectionnels à étudier.

En plus du choix des attributs sensibles, les effectifs des groupes intersectionnels représentent un point important de la fiabilité des mesures. En effet, à partir de l'importance de l'analyse par groupes intersectionnels, rappelée dans la section précédente, la question qui vient à se poser pour aller plus loin est : comment détecter toutes les discriminations possibles? Autrement dit, comment déterminer le niveau d'intersectionnalité des groupes à considérer? En effet, le niveau maximal d'intersectionnalité serait de considérer tous les attributs (sensibles ou non) d'un jeu de données s'ils sont catégoriels, ou de comparer les résultats des individus deux à deux s'il y a au moins un attribut numérique. Cependant, même si cela permet une recherche systématique des discriminations, en testant tous les attributs, il semble difficile de remonter au sens des discriminations quand les groupes intersectionnels sont trop fins, c'est-à-dire à l'intersection d'un trop grand nombre d'attributs, et par conséquent regroupant un trop faible nombre d'individus. Ce point sur les effectifs est en effet une perspective intéressante à examiner en lien avec l'équité : à partir de quelles proportions d'individus dans un jeu de données est-il pertinent de les comparer en termes d'équité algorithmique?

7.5. RECOMMANDATIONS PRATIQUES

Pour finir, nous proposons des recommandations pratiques pour utiliser la MADD à destination des chercheuses, chercheurs, développeuses et développeurs de modèles prédictifs en éducation. Tout d'abord, nous mettons à disposition une librairie Python, `maddlib` (lien à la note de pied de page n°4), pour partager des outils communs à l'analyse de l'équité. Ensuite, nous proposons les instructions suivantes, en 7 étapes, pour conduire une analyse d'équité, que ce soit avec des attributs individuels ou des groupes intersectionnels :

1. Choisir des modèles de classification binaire qui, en plus de leurs prédictions, produisent des estimations de probabilité ou des scores de confiance.
2. Transformer, si besoin, les attributs sensibles numériques en attributs binaires ou catégoriels.
3. Entraîner les modèles choisis.
4. En phase de test, séparer les probabilités prédites en fonction de chaque groupe à considérer, formé à partir d'un ou plusieurs attributs sensibles.
5. Calculer la MADD entre deux groupes d'intérêt (voir Section 6). Répéter le calcul autant de fois que de comparaison souhaitée.
6. Visualiser le comportement des modèles vis-à-vis des deux groupes d'intérêt et identifier graphiquement un traitement inégal ou un jugement stéréotypé des modèles (voir Verger, Bouchet *et al.* (2023) pour leurs explications).
7. Analyser les discriminations algorithmiques selon l'objectif recherché, dont voici des exemples principaux :
 - Identifier le modèle le plus équitable en moyenne ou par attribut, par une lecture en ligne des tableaux de résultats (voir Section 6).
 - Identifier l'attribut le plus équitable en moyenne ou par modèle, par une lecture en colonne des tableaux de résultats (voir Section 6).
 - Identifier le groupe le plus discriminé, par analyse visuelle (Verger, Bouchet *et al.*, 2023), par un classement (voir Figure 5) ou une comparaison deux à deux (voir Figure 6).

Les étapes n°4 à 7 sont notamment facilitées par les outils de la librairie `maddlib`.

8. CONCLUSION

Dans cet article, nous avons conduit des analyses approfondies de l'équité algorithmique grâce à l'approche intersectionnelle. Nous avons mis en évidence les différences détectées selon les attributs sensibles individuels et les groupes intersectionnels présents dans le contexte éducatif du OULAD (Kuzilek *et al.*, 2017). Ces analyses montrent l'importance d'inclure une approche intersectionnelle pour l'évaluation de l'équité de tout modèle prédictif, afin de détecter et de mieux comprendre certaines discriminations, au-delà des analyses traditionnellement menées sur les attributs sensibles seuls (voir Section 2).

Les deux principaux résultats de ces analyses sont les suivants. D'une part, les analyses ont permis de mettre en lumière des discriminations qu'on ne pouvait voir avec l'analyse des attributs individuels seuls. D'autre part, elles ont permis de comprendre de manière plus fine l'influence de chaque attribut sur les discriminations grâce à leurs interactions avec les autres attributs. Ces nouveaux résultats répondent à la QR2 posées en Section 1, grâce aux approches que nous avons développées pour la QR1, et ils renforcent la position de Verger, Bouchet *et al.* (2023) sur la nécessité d'analyser systématiquement les discriminations algorithmiques des modèles prédictifs pour chaque application éducative. Elles permettent en effet de fournir des éclairages sur les implications réelles de l'utilisation de ces modèles ainsi que sur les discriminations qui pourraient exister.

De plus, dans un souci de reproductibilité des résultats et d'utilisation de notre méthode dans d'autres contextes éducatifs, nous mettons à disposition les données et le code documenté⁶ ainsi que la librairie Python que nous avons nommé `maddlib`⁷.

6. <https://github.com/melinaverger/MADD>

7. <https://pypi.org/project/maddlib>

Enfin, les perspectives pourront s'orienter vers l'étude de l'influence des effectifs des groupes intersectionnels sur les mesures d'équité, qui est un point crucial pour la fiabilité de ces mesures. Il est également intéressant d'explorer des méthodes d'atténuation des discriminations algorithmiques avec une approche intersectionnelle plutôt que de réduire ces discriminations par rapport à un seul attribut. Notamment, nous avons proposé une méthode de post-traitement utilisant la MADD (Verger, Fan *et al.*, 2023 ; Verger *et al.*, 2024) qui pourrait conduire à une expérience avec des groupes intersectionnels. Néanmoins, bien que toute méthode d'atténuation des discriminations algorithmiques (Caton et Haas, 2020 ; Mehrabi *et al.*, 2022) offre une première piste vers l'utilisation de systèmes plus équitables, elle ne prémunit pas contre une évaluation pas seulement informatique mais plus globale de l'équité des systèmes, avec la concertation des institutions, des enseignantes et enseignants, et des apprenantes et apprenants.

RÉFÉRENCES

- Alhakbani, H. A., et Alnassar, F. M. (2022). Open learning analytics : a systematic review of benchmark studies using Open University learning analytics dataset (OULAD). *7th International Conference on Machine Learning Technologies (ICMLT)*, (p. 81-86). <https://doi.org/10.1145/3529399.3529413>
- Baker, R. S., et Hawn, A. (2022). Algorithmic bias in education. *International Journal of Artificial Intelligence in Education*, 18, 1052-1092. <https://doi.org/10.1007/s40593-021-00285-9>
- Barocas, S., Hardt, M., et Narayanan, A. (2019). *Fairness and machine learning : limitations and opportunities* [<http://www.fairmlbook.org>]. MIT Press.
- Buolamwini, J., et Gebru, T. (2018). Gender shades : intersectional accuracy disparities in commercial gender classification. *1st Conference on Fairness, Accountability and Transparency*, (p. 77-91). <https://proceedings.mlr.press/v81/buolamwini18a.html>
- Castelnovo, A., Crupi, R., Greco, G., Regoli, D., Penco, I. G., et Cosentini, A. C. (2022). A clarification of the nuances in the fairness metrics landscape. *Scientific Reports*, 12(4209). <https://doi.org/10.1038/s41598-022-07939-1>
- Caton, S., et Haas, C. (2020). Fairness in machine learning : a survey. *ACM Computing Surveys*, 56(7), 1-38. <https://api.semanticscholar.org/CorpusID:222208640>
- Crenshaw, K. (1989). Demarginalizing the intersection of race and sex : a black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics. *The University of Chicago Legal Forum*, 140, 139-167.
- Deho, O. B., Zhan, C., Li, J., Liu, J., Liu, L., et Duy Le, T. (2022). How do the existing fairness metrics and unfairness mitigation algorithms contribute to ethical learning analytics ? *British Journal of Educational Technology*, 53(4), 822-843. <https://doi.org/https://doi.org/10.1111/bjet.13217>
- Evans-Winters, V. E. (2021). Race and gender intersectionality and education. *Oxford Research Encyclopedia of Education*, 42, 1-27.
- Gardner, J., Brooks, C., et Baker, R. (2019). Evaluating the fairness of predictive student models through slicing analysis. *9th International Conference on Learning Analytics & Knowledge*, (p. 225-234). <https://doi.org/10.1145/3303772.3303791>
- Gohar, U., et Cheng, L. (2023). A survey on intersectional fairness in machine learning : notions, mitigation, and challenges. *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*, 6619-6627. <https://doi.org/10.24963/ijcai.2023/742>

- Hellas, A., Ihanola, P., Petersen, A., Ajanovski, V. V., Gutica, M., Hynninen, T., Knutas, A., Leinonen, J., Messom, C., et Liao, S. N. (2018). Predicting academic performance : a systematic literature review. *23rd Annual ACM Conference on Innovation and Technology in Computer Science Education*, (p. 175-199). <https://doi.org/10.1145/3293881.3295783>
- Hu, Q., et Rangwala, H. (2020). Towards fair educational data mining : a case study on detecting at-risk students. *13th International Conference on Educational Data Mining*, (p. 7).
- Kilbertus, N., Rojas-Carulla, M., Parascandolo, G., Hardt, M., Janzing, D., et Schölkopf, B. (2017). Avoiding discrimination through causal reasoning. *Proceedings of the 31st International Conference on Neural Information Processing Systems*, (p. 656-666).
- Kizilcec, R. F., et Lee, H. (2022). Algorithmic fairness in education. Dans *The ethics of artificial intelligence in education* (p. 174-202). Routledge.
- Korkmaz, C., et Correia, A.-P. (2019). A review of research on machine learning in educational technology. *Educational Media International*, 56(3), 250-267. <https://doi.org/10.1080/09523987.2019.1669875>
- Kullback, S. (1959). *Information theory and statistics*. Wiley.
- Kuzilek, J., Hlosta, M., et Zdrahal, Z. (2017). Open University learning analytics dataset. *Sci Data*, 4(1), 1-8. <https://doi.org/https://doi.org/10.1038/sdata.2017.171>
- Lee, H., et Kizilcec, R. F. (2020). Evaluation of fairness trade-offs in predicting student success. *arXiv preprint arXiv :2007.00088*. <http://arxiv.org/abs/2007.00088>
- Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., et Galstyan, A. (2022). A survey on bias and fairness in machine learning. *arXiv :1908.09635 [cs]*. <http://arxiv.org/abs/1908.09635>
- Pearl, J. (2009). *Causality* (2^e éd.). Cambridge University Press.
- Pessach, D., et Shmueli, E. (2022). A review on fairness in machine learning. *ACM Computing Surveys*, 55(3), 1-44. <https://doi.org/10.1145/3494672>
- Romero, C., et Ventura, S. (2020). Educational data mining and learning analytics : an updated survey. *Wiley Interdisciplinary Reviews : Data Mining and Knowledge Discovery*, 10(3), e1355. <https://doi.org/10.1002/widm.1355>
- Verger, M. (2022). Investiguer la notion d'équité algorithmique dans les environnements informatiques pour l'apprentissage humain. *9ièmes RJC EIAH 2022 : Environnements Informatiques pour l'Apprentissage Humain*.
- Verger, M., Bouchet, F., Lallé, S., et Luengo, V. (2023). Caractérisation et mesure des discriminations algorithmiques dans la prédiction de la réussite à des cours en ligne. *11ème Conférence sur les Environnements Informatiques pour l'Apprentissage Humain (EIAH 2023)*.
- Verger, M., Fan, C., Lallé, S., Bouchet, F., et Luengo, V. (2023). A fair post-processing method based on the MADD metric for predictive student models. *st International Tutorial and Workshop on Responsible Knowledge Discovery in Education (RKDE 2023) at ECML PKDD 2023, Turino, Italy*. hal-04345451v1
- Verger, M., Fan, C., Lallé, S., Bouchet, F., et Luengo, V. (2024). A comprehensive study on evaluating and mitigating algorithmic unfairness with the MADD metric. *Journal of Educational Data Mining (JEDM)*, 16(1), 365-409.
- Verger, M., Lallé, S., Bouchet, F., et Luengo, V. (2023). Is your model "MADD"? A novel metric to evaluate algorithmic fairness for predictive student models. *16th International Conference on Educational Data Mining*.
- Verma, S., et Rubin, J. S. (2018). Fairness definitions explained. *2018 IEEE/ACM International Workshop on Software Fairness (FairWare)*, (p. 1-7).

- Yang, F., Cisse, M., et Koyejo, S. (2020). Fairness with overlapping groups. *Advances in Neural Information Processing Systems*, 33.
- Zambrano, A. F., Zhang, J., et Baker, R. S. (2024). Investigating algorithmic bias on bayesian knowledge tracing and carelessness detectors. *Proceedings of the 14th Learning Analytics and Knowledge Conference*, (p. 349-359).

Adaptive Learning en contexte parascolaire : comprendre les usages et effets *via* l'analyse des traces d'un déploiement industriel

Adaptive Learning in Out-of-School Context : Understanding Uses and Effects by Analyzing the Traces of an Industrial Deployment

Anaëlle BADIER¹; Marie LEFEVRE¹; Mathieu LEFORT¹; Nathalie GUIN¹

¹Université Claude Bernard Lyon 1, CNRS, École Centrale de Lyon, INSA Lyon, Université Lumière Lyon 2, LIRIS, UMR5205, 69622 Villeurbanne, France

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, s'appuyant sur l'IRT (Item Response Theory) et sur un score de recommandation à trois composantes : pédagogique, historique et nouveauté. 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 8 000 apprenants sur 4 mois. Nous montrons que les recommandations répondent à des objectifs et des usages différents, que leur suivi influe positivement sur l'expérience d'apprentissage et que l'analyse de ces traces permet de mettre en évidence des leviers d'amélioration pour proposer un mécanisme de recommandation par conception itérative.

Mots-clés : adaptive learning, système de recommandations, traces d'apprentissage, apprentissage non-formel.

Abstract. *Our work takes place in a context of non-formal learning on mobile applications, where we have proposed a recommendation engine for educational resources, based on IRT (Item Response Theory) and a recommendation score with three components: pedagogical, historical and novelty. The learners concerned have a wide range of objectives and work methods. In this paper, after explaining how our recommendation engine works, we analyze the usage traces of over 8,000 learners over a 4-month period. We show that the recommendations meet different objectives and uses, that following them has a positive impact on the learning experience, and that analysis of these traces highlights levers for improvement in proposing a recommendation mechanism based on iterative design.*

Keywords: *adaptive learning, recommender system, learning traces, out-of-school.*

1. INTRODUCTION

Le nombre croissant des applications et plateformes de e-learning depuis plusieurs décennies a conduit à de nouvelles formes d'apprentissage. En parallèle, les conditions d'enseignement en classe dans l'Éducation Nationale (créneaux alloués à chaque matière dans l'emploi du temps, effectif des groupes) ne permettent pas toujours de s'adapter aux besoins des élèves. Certains d'entre eux recherchent donc dans les outils de e-learning, un complément à la formation qu'ils reçoivent en classe. Ils sont alors amenés à travailler par eux-mêmes en recherchant des contenus d'apprentissage hors de la classe. Il peut dans ce cas s'avérer difficile pour un apprenant d'identifier ses besoins d'apprentissage et de sélectionner les ressources adaptées, d'autant plus si ces ressources sont à extraire de programmes scolaires distincts de son cursus actuel, en réponse à d'éventuelles lacunes héritées des années antérieures ou pour susciter une curiosité pour l'étude de chapitres qui interviendront plus tard dans sa scolarité.

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 d'apprentissage. 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 (Badier *et al.*, 2023).

Dans cet article, nous nous intéressons à l'expérience d'apprentissage avec ce système au travers les questions de recherche suivantes :

QR1- Quels facteurs influencent le suivi des recommandations par les apprenants dans un contexte parascolaire ?

QR2 - Quels usages des recommandations pouvons-nous observer dans une application parascolaire ?

QR3 - Peut-on mettre en évidence des leviers d'amélioration du modèle de recommandations à partir des analyses de traces issues de la navigation libre des apprenants sur l'application ?

Après avoir présenté l'application mobile avec laquelle nous travaillons (section 2), nous détaillons le modèle et le principe de fonctionnement du système de recommandations implémenté dans l'application mobile, en le positionnant au regard des travaux du domaine (section 3). Nous présentons ensuite le cadre d'expérimentation et de recueil des données d'utilisation (section 4). Nous exposons enfin notre étude exploratoire : nous analysons selon plusieurs critères les recommandations suivies par les apprenants parmi celles proposées (section 5), nous étudions les modifications observées dans l'expérience d'apprentissage (section 6), puis nous cherchons à comprendre les motifs de non suivi des recommandations (section 7). Les résultats présentés sont discutés (section 8), avant de terminer sur les perspectives ouvertes par ces travaux (section 9).

2. APPLICATION MOBILE PARASCOLAIRE

Nos travaux sont conduits au sein de l'application mobile Nomad Education. Par conséquent, pour proposer un système de recommandations pour cette application, nous avons dû prendre en compte l'existant. Cette section a pour objectif de présenter l'application, ainsi

que les ressources pédagogiques dont nous disposons et leur organisation au sein de l'application.

2.1. PRÉSENTATION DE L'APPLICATION MOBILE

Nomad Education est une application mobile disponible sur smartphones et tablettes qui propose les contenus pédagogiques du programme scolaire français, rédigés par des professeurs de l'Éducation Nationale. Cette application s'adresse aux apprenants des différents niveaux scolaires, du CP jusqu'au supérieur. Dans le cadre de cette étude, nous restreignons l'étude sur les niveaux Collège-Lycée car ils constituent la population d'utilisateurs majoritaire de l'application.

L'application est destinée à un **usage parascolaire, volontaire et non supervisé**. Elle n'est pas intégrée dans le parcours scolaire principal de l'élève et n'est pas pensée pour établir un lien entre les contenus étudiés par l'élève au sein de son établissement scolaire, ses notes ou ses professeurs, et le contenu étudié au sein de l'application. On se situe donc dans un contexte d'apprentissage parascolaire qualifié de *mobile-learning*, *micro-learning* et *non formel*.

Des captures d'écran de l'interface de l'application sont présentées sur la Figure 1. Au téléchargement de l'application, les apprenants remplissent leur profil en renseignant notamment leur classe, les options choisies dans le cadre de leur cursus scolaire et leurs souhaits d'orientation pour l'année suivante. Après une première synchronisation des contenus, ils accèdent à la page d'accueil (voir écran de gauche de la Figure 1). Sur cette page, ils peuvent visualiser leur temps de travail hebdomadaire, accéder à quelques fonctionnalités de gamification ainsi qu'à certaines rubriques de contenus hors programme scolaire. Ces rubriques présentent des sujets d'actualité, de sport ou de citoyenneté (Parcoursup, Jeux Olympiques, explication sur les élections...). À ce jour, l'utilisation de l'application est strictement individuelle, il n'y a pas de fonctionnalité d'apprentissage collaboratif. Dans la rubrique *Révisions* (voir écran au centre de la Figure 1), les apprenants ont accès à toutes les disciplines du programme scolaire de leur année en cours. Occasionnellement, selon certains niveaux scolaires, des contenus annexes leur sont proposés, comme des quiz ou des fiches méthodologies pour aider à la préparation des épreuves nationales. La partie droite de la Figure 1 présente un aperçu des contenus accessibles au sein d'une discipline. Les ressources disponibles dans cette rubrique seront détaillées dans la section 2.2.

Les ressources pédagogiques de l'application mobile Nomad Education sont conçues par des professeurs de l'Éducation Nationale en respectant le programme officiel. Toutefois, l'organisation de ce programme diffère selon les niveaux scolaires, et la distinction entre compétence, connaissance et savoir-faire n'y est pas explicite. Dans le Bulletin Officiel de l'Éducation Nationale pour le cycle 3 correspondant aux niveaux CM1-CM2-6^{ème} (BO, 2020), il est évoqué à la fois des "attendus de fin de cycle", des "compétences travaillées", "compétences et connaissances associées" sans distinction faite entre les unes et les autres, ainsi que des "exemples de situations, d'activités et outils pour l'élève". Dans le programme officiel de Mathématiques en 1^{ère} générale (2022), le texte reprend les 6 compétences mathématiques essentielles (chercher, modéliser, représenter, raisonner, calculer, communiquer) correspondantes aux "compétences travaillées" du bulletin pour le 3^{ème} cycle mais le bulletin est ensuite structuré en "contenus d'enseignement", lesquels sont divisés en "situations et problèmes" faisant appel à des "contenus mathématiques", pour lesquels il y a des "capacités attendues". Ces bulletins ne permettant pas d'extraire directement des connaissances, composantes ou savoir-faire pour les associer aux ressources, au sein de l'application Nomad Education, les ressources sont donc décrites en fonction des savoirs mobilisés. Cette description est détaillée dans la section 2.3.

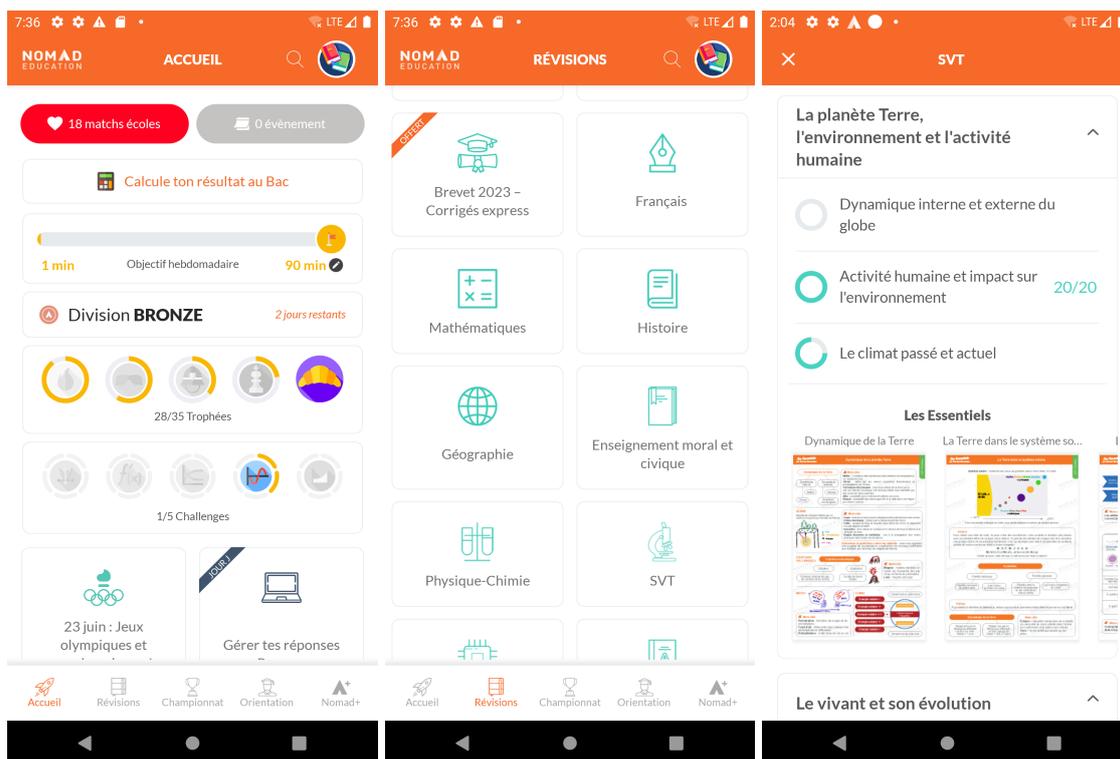


Figure 1 : Captures d'écran de l'application Nomad Education : à gauche la page "Accueil" commune à tous les niveaux, au milieu la page "Révisions" pour un niveau 3^{ème}, à droite un aperçu de la discipline SVT pour un niveau 3^{ème}

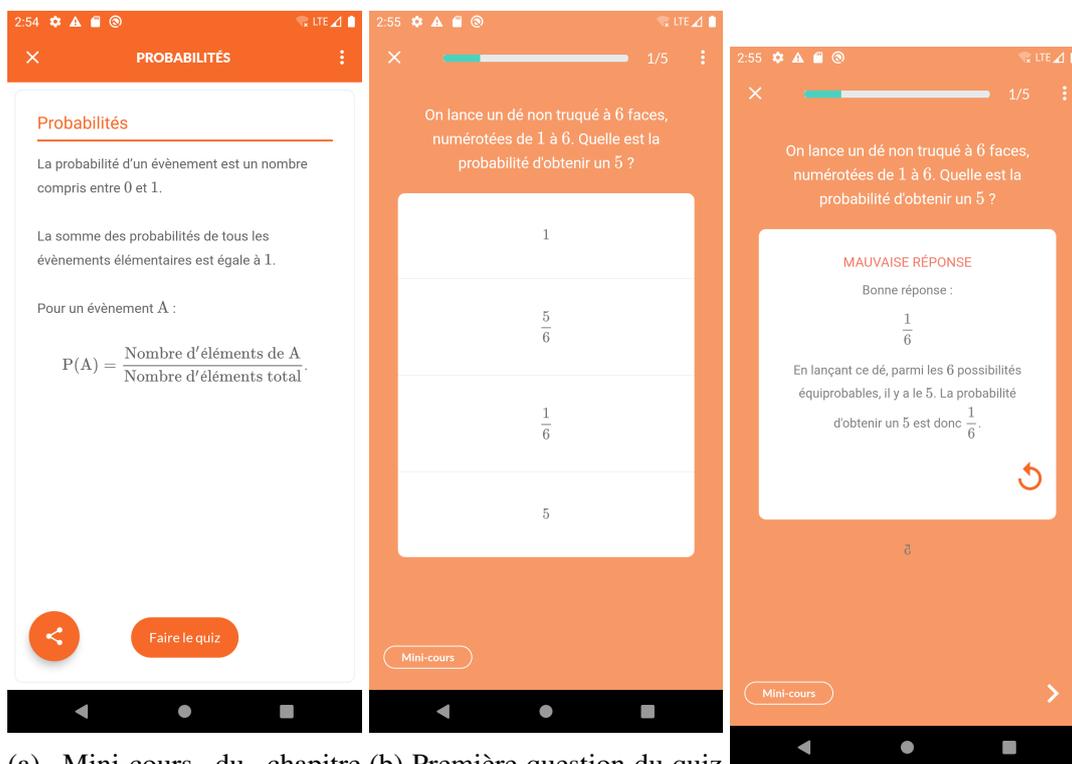
2.2. RESSOURCES DISPONIBLES DANS L'APPLICATION NOMAD EDUCATION

Pour chaque programme scolaire (6^{ème}, 2^{nde}, Terminale...), l'application est organisée en disciplines (Mathématiques, Français, Sciences de la Vie et de la Terre (SVT)...), elles-mêmes subdivisées en chapitres (*Probabilités*, *Volumes*, *Fonction linéaire*, ... en Mathématiques; *La nutrition des organismes*, *Le climat passé et actuel*, ... en SVT; etc). Chaque chapitre contient de 1 à 5 mini-cours (moins de 200 mots) accessibles en lecture ou en podcast, et 4 quiz de 5 questions au format QCM. Pour certaines disciplines, des fiches de révision au format PDF sont parfois ajoutées. Un exemple de mini-cours et de question de quiz pour le chapitre *Probabilités* de la discipline Mathématiques de niveau 5^{ème} sont visibles sur la Figure 2. L'apprenant a la possibilité de recommencer ses quiz autant de fois qu'il le souhaite selon trois modalités : affichage des questions dans le même ordre, dans le désordre ou recommencer uniquement les questions pour lesquelles les réponses apportées étaient fausses. Pendant la réalisation du quiz, l'apprenant peut à tout moment consulter le mini-cours du chapitre correspondant (bouton en bas à gauche de l'écran).

L'application mobile contient donc 6 types d'objets d'apprentissage : *Programme scolaire*, *Discipline*, *Chapitre*, *Quiz*, *Question*, *Cours*. Ces objets sont stockés au format JSON dans une base de données MongoDB. Pour chaque ressource sont renseignés le type de la ressource (chapitre, quiz, cours...), son titre, le contenu, et l'entité de rattachement (un quiz est rattaché à un chapitre, un chapitre est rattaché à une discipline...). Aucune méta-donnée descriptive supplémentaire n'est renseignée.

Au sein de chaque discipline, les chapitres sont rédigés pour respecter le programme de l'Éducation Nationale. Toutefois, les contenus étant exclusivement ceux produits pour l'application Nomad Education, le nombre de ressources à disposition pour aborder une

thématique donnée est relativement restreint. Le tableau 1 présente le nombre de chapitres disponibles dans l'application pour deux disciplines, Mathématiques et SVT, sur les niveaux Collège et Lycée. Comme le montre ce tableau, l'ensemble du programme de Mathématiques de 4^{ème} est contenu dans 15 chapitres, chacun ne contenant que 4 quiz de 5 questions chacun. De ce fait, pour un niveau scolaire donné, un même sujet (par exemple les *probabilités*) n'est traité que par un seul chapitre.



(a) Mini-cours du chapitre *Probabilités* pour le niveau 5^{ème} (b) Première question du quiz *Probabilités* pour le niveau 5^{ème} (c) Correction proposée pour la question visible figure 2b

Figure 2 : Exemple d'objets d'apprentissage dans l'application mobile Nomad Education

Tableau 1 : Nombre de chapitres disponibles pour les niveaux scolaires Collège-Lycée dans l'application mobile Nomad Education

Niveau scolaire	Chapitres en Mathématiques	Chapitres en SVT
6e	12	8
5e	15	9
4e	15	9
3e	17	9
2nde	10	17
1ère Générale	10	16
Terminale Générale	23 (+24 en option maths)	20

2.3. UN GRAPHE DE NOTIONS POUR DÉCRIRE LES RESSOURCES

Partant du constat que les programmes scolaires suivent une continuité dans les concepts abordés, les professeurs-concepteurs des chapitres de l'application ont associé à chacun des concepts les notions concernées par le chapitre en question.

Dans l'application Nomad Education, une **notion** est définie comme une "composante de savoir abordée dans une ressource pédagogique donnée". Le terme de "notion" est préféré aux termes "connaissances" ou "compétences" car bien qu'inspirés du "socle commun de connaissances, de compétences et de culture" (Décret, 2015) adopté par l'Éducation Nationale, ces mots-clés apposés sur les chapitres ne sont pas uniquement extraits des programmes scolaires et peuvent être reformulés. Le programme de l'Éducation Nationale utilise la forme infinitive pour citer les compétences (*chercher, modéliser, représenter*). Les notions apposées sur les chapitres de l'application décrivent des thématiques plus précises sous la forme de noms communs (*Volume, Angle, Nombres décimaux*), davantage assimilables aux "contenus mathématiques" listés dans le bulletin de l'Éducation Nationale (*Fonctions affines, Suites arithmétiques, Tableau de variation, etc.*) (BO, 2020).

Sur chaque chapitre, les notions apposées sont qualifiées de "**type prérequis**" s'il s'agit d'une composante de connaissance élémentaire dans le chapitre ou de "**type attendu**" si la notion est découverte ou sera davantage maîtrisée suite à l'étude du chapitre. À noter que comme il n'est pas question de compétences, cette qualification en *prérequis/attendu* induit une relation d'ordre moins stricte que dans les travaux de De Maio *et al.* (2012) ou El-Kechai *et al.* (2015). Une notion donnée peut-être qualifiée de prérequis et d'attendue sur un même chapitre si ce chapitre permet de monter en compétence sur cette notion.

Prenons par exemple le chapitre *Probabilités* de niveau 5^{ème} présenté dans la section précédente (voir écran de gauche sur la Figure 2). Sur ce chapitre sont apposés 5 notions prérequis : *Fraction, Quotient, Fréquence, Pourcentages, Addition*. En effet, les probabilités sont abordées dans le programme scolaire de Mathématiques pour la première fois en classe de 5^{ème}. Les notions prérequis font donc appel à des notions d'arithmétique de base, notamment à la maîtrise des fractions comme en témoigne la formule de calcul d'une probabilité (voir Figure 2a). Sur ce même chapitre, 5 notions ont été apposées comme attendues : *Probabilités, Equiprobabilité, Fraction, Quotient, Fréquence*. Les notions *Probabilités* et *Equiprobabilité* apparaissent en notions attendues car elles sont découvertes grâce à l'étude de ce chapitre. Les notions *Fraction, Quotient, Fréquence* étaient déjà des notions prérequis qui apparaissent, à nouveau, en notions attendues car ce chapitre permet également à l'élève de travailler ces notions et donc d'améliorer son niveau sur ces concepts.

Le terme de **tag** permet de nommer le triplet : $\langle \text{chapitre, notion, type} \rangle$. Pour ce même chapitre *Probabilités* de niveau 5^{ème}, nous avons donc 10 tags : [$\langle \text{Probabilités (5ème), Fraction, Prérequis} \rangle, \langle \text{Probabilités (5ème), Quotient, Prérequis} \rangle, \langle \text{Probabilités (5ème), Fraction, Attendu} \rangle, \dots$].

Des tags sont ainsi apposés par les professeurs sur chaque chapitre, établissant ainsi des liens entre deux chapitres annotés avec des notions similaires. Cette association de notions sur les chapitres constitue un **graphe de notions**, dont un exemple partiel est représenté sur la Figure 3. Sur cet extrait, le chapitre de *Probabilité 5^{ème}* est lié au chapitre *Proportionnalité 6^{ème}* car ils ont en commun la notion de *Pourcentage*, au chapitre de *Probabilité 4^{ème}* car ils ont en commun les notions *Probabilités, Fréquence* et au chapitre *Probabilités conditionnelles 1^{ère}* car ils ont en commun la notion *Probabilités*. Ce graphe permet de visualiser les connexions entre les différents sujets abordés dans le programme d'études.

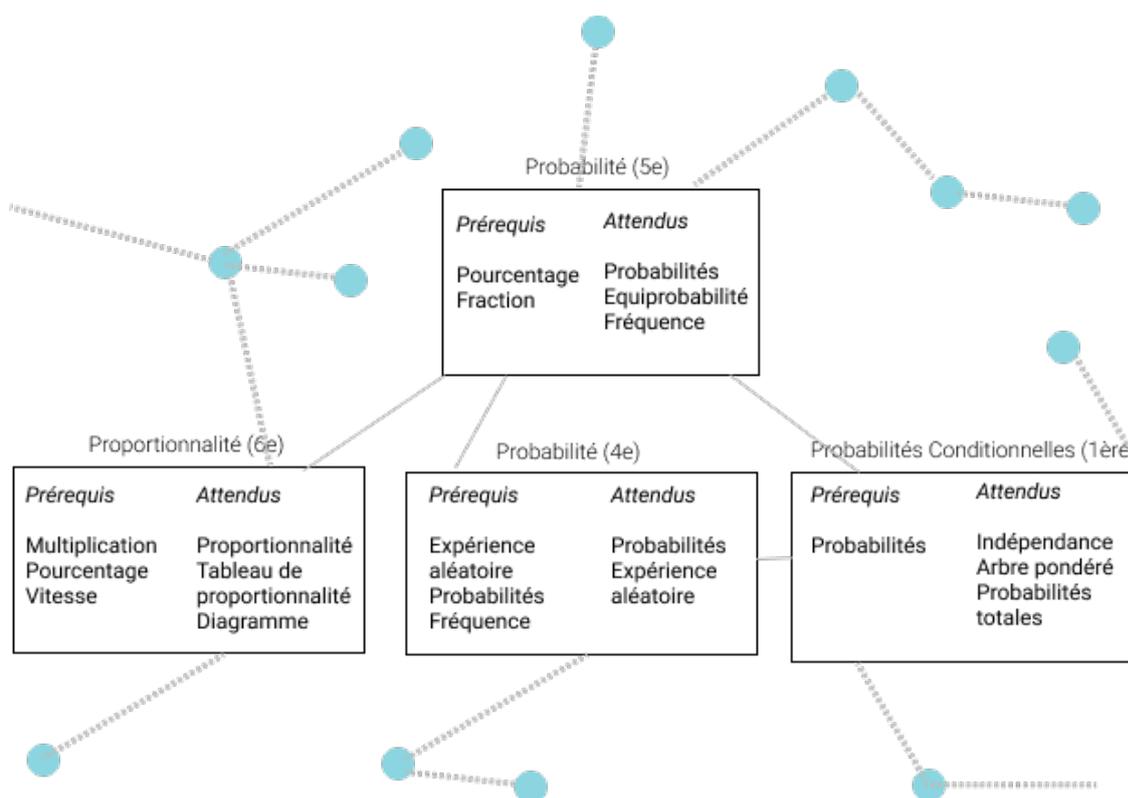


Figure 3 : Extrait du graphe de notions de l'application Nomad Education pour la discipline Mathématiques montrant un focus sur le chapitre *Probabilité 5^{ème}* et ses liens avec les autres chapitres de l'application, et illustrant en arrière plan, que ces autres chapitres ont également d'autres connexions

3. RECOMMANDER DANS UNE APPLICATION MOBILE PARASCOLAIRE

Comme nous travaillons avec une application mobile, notre contexte est assez similaire à celui des plateformes MOOC (Reich, 2014), mais nos apprenants utilisent cette application comme un complément et non comme leur principal support d'apprentissage. L'application peut être définie comme un outil de micro-apprentissage (Nikou et Economides, 2018).

Selon les revues systématiques des systèmes de recommandations dans l'éducation (Urdaneta-Ponte *et al.*, 2021), les stratégies de recommandations sont essentiellement basées sur le contenu et sur l'apprenant (y compris les stratégies hybrides). Plusieurs méthodes sont utilisées, comme le filtrage collaboratif, le filtrage basé sur le contenu, ou les techniques de *Data Mining* (Guruge *et al.*, 2021).

Pour fournir des recommandations et s'adapter aux utilisateurs, certains systèmes sont basés sur le concept de "Zone de Développement Proximal" (ZPD) développé par Vygotskiï et Cole (1978). Elle consiste à évaluer la zone où l'élève est capable d'exécuter une tâche. Cette ZPD peut être utilisée pour recommander à l'élève un contenu d'apprentissage un peu plus difficile, afin de le faire progresser sans le frustrer avec des ressources trop faciles ou trop difficiles. Baker *et al.* (2020) ont utilisé un système de recommandations basé sur la ZPD et ont prouvé ses effets positifs sur l'apprentissage. Cette méthode nécessite d'évaluer le niveau de compétences de l'étudiant. Plusieurs systèmes utilisent la taxonomie des compétences de Bloom (1956) pour s'adapter aux compétences de l'apprenant. Les modèles de traçage des connaissances (Corbett et Anderson, 1994) sont largement utilisés pour déduire les compétences et modéliser le niveau de connaissance d'un étudiant (Vie et Kashima,

2019). Ces modèles visent à prédire les résultats des étudiants en fonction des questions. À l'aide de statistiques, la théorie basée sur les items (IRT) (Baker, 2001) est une méthode utilisée pour évaluer le niveau latent d'un individu (par exemple son niveau de compétence) et la probabilité de réussir un item. L'utilisation de l'IRT dans les EIAH a été étudiée par Wauters *et al.* (2021).

Les systèmes de recommandations peuvent être des systèmes de "soutien à la résolution de problèmes" (Brusilovsky et Peylo, 2003), développés pour aider les étudiants à résoudre une tâche précise dans un domaine particulier, par exemple pour apprendre l'arithmétique (Clément *et al.*, 2015) ou les langages de programmation (Branthôme, 2022). Dans l'application Nomad Education, le système de recommandation est utilisé pour différentes disciplines, c'est pourquoi les travaux en didactique ne sont pas exploités. D'autres EIAH sont basés sur des ressources en ligne (Daher *et al.*, 2018). Le contenu est structuré, généralement avec des ontologies (Murray, 2003 ; Nguyen *et al.*, 2014), et peut être organisé en graphes de connaissances (Rizun, 2019). Le système de recommandation de l'application Nomad Education doit être adapté à plusieurs disciplines, et il ne dispose pas d'autant de ressources que les systèmes basés sur le web (De Maio *et al.*, 2012). Cependant, il exploite certaines méthodes précédemment décrites comme les stratégies basées sur le contenu et l'IRT. Dans notre système de recommandation, implémenté dans l'application Nomad Education, l'IRT est détourné de son utilisation classique, et sert à deux objectifs : évaluer les connaissances d'un étudiant mais aussi décrire le niveau de difficulté des ressources pédagogiques.

Le modèle que nous présentons dans la section suivante est conçu pour gérer une utilisation volontaire, irrégulière et autonome avec des ressources internes limitées. En raison de notre contexte, les méthodes existantes ne sont pas directement applicables. Par conséquent, certains critères d'évaluation des systèmes de recommandations (Erdt *et al.*, 2015), comme la mesure des effets sur l'apprentissage, ne peuvent pas non plus être utilisés. En effet, cette mesure ne peut être calculée que si nous maîtrisons le cycle d'apprentissage complet et que nous sommes donc sûrs que le gain d'apprentissage est dû à l'application proposée. Or, comme nous travaillons avec une application extrascolaire, l'apprentissage se fait à travers l'application et en dehors de l'application. Par conséquent, nous n'utiliserons que les critères d'évaluation des systèmes de recommandations qui s'appliquent à notre contexte d'apprentissage libre, volontaire et complémentaire à l'apprentissage scolaire (voir section 4).

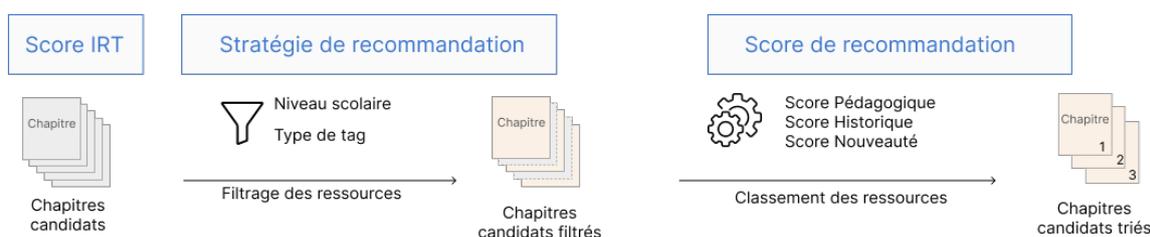


Figure 4 : Principe du modèle de recommandations pour une application mobile parascolaire

Notre modèle de recommandation est résumé sur la Figure 4 et détaillé dans les sections suivantes. Dans un premier temps, nous exploitons l'approche IRT pour définir 3 niveaux de maîtrise d'un chapitre et assigner à chaque niveau de maîtrise une stratégie de recommandation : révision pour les apprenants de niveau inférieur, continuité pour les apprenants de niveau intermédiaire et approfondissement pour les apprenants de niveau supérieur. Pour chaque stratégie, nous mobilisons les tags apposés sur les chapitres de l'application d'après le graphe de notions de la discipline concernée, pour sélectionner les chapitres liés aux notions prérequis ou attendues du chapitre en cours d'étude. Ensuite, un score de recommandation

nous permet de classer par pertinence les chapitres pré-sélectionnés selon la stratégie. Nous avons 3 composantes scorées pour quantifier la pertinence d'un chapitre par rapport à un autre dans un contexte et pour un apprenant donné. La composante *Pédagogique* permet de s'assurer que le chapitre recommandé est proche en terme de notions communes avec le chapitre en cours d'étude et en termes d'écart au niveau scolaire. La composante *Historique* nous permet de prendre en compte l'historique de l'apprenant pour recommander des contenus qui soient en cohérence avec les chapitres qu'il a précédemment étudié sur l'application et les résultats obtenus sur ces chapitres. La composante *Nouveauté* permet d'inclure de la variété dans les recommandations proposées, afin de s'adapter au contexte du *mobile-learning* avec des apprenants qui peuvent espérer des recommandations diversifiées et renouvelées.

3.1. EXPLOITER L'IRT POUR ESTIMER LE NIVEAU DE L'APPRENANT

L'IRT est classiquement utilisée pour concevoir des tests adaptatifs administrés par ordinateur, dont le but est d'estimer au plus près le niveau de capacité réel de l'apprenant avec un minimum de questions (Vie, 2016). Dans l'application Nomad Education, l'approche IRT a été étudiée non pas pour concevoir des tests adaptatifs mais pour avoir une estimation des niveaux de capacité réels des apprenants, point d'entrée de notre système de recommandations. Dans la littérature, on relève également d'autres travaux qui utilisent l'IRT dans le cadre de systèmes de recommandations afin d'estimer les niveaux de capacité des étudiants (Ortegón Romero et Krug Wives, 2021). On privilégie l'approche par IRT par rapport à la moyenne des notes brutes pour définir ce niveau de capacité car l'IRT répond à deux potentiels problèmes identifiés si l'on considérait simplement la note obtenue sur les quiz. D'une part, l'IRT permet de considérer grâce au paramètre b la difficulté des items, ce qui est essentiel dans notre contexte car ce sont des professeurs différents qui rédigent les quiz, ceux-ci n'ont donc pas un niveau de difficulté standardisé. D'autre part, nous pensons que le contexte d'apprentissage auto-régulé sans incidence des notes obtenues au sein de l'application sur les dossiers scolaires des élèves peut inciter ces derniers à répondre davantage au hasard aux questions posées. C'est par ailleurs une possibilité d'apprentissage de fonctionner par essais-erreurs en répondant au hasard à la question et en lisant la correction associée (Pinsky et Irby, 1997). Par conséquent, la possibilité de trouver la bonne réponse en répondant au hasard n'est pas à négliger, ce qui est pris en compte par le paramètre c du modèle IRT à 3 paramètres (Baker, 2001) représenté par l'équation 1 :

$$P(\theta) = c + (1 - c) \frac{1}{1 + e^{-a(\theta - b)}} \quad (1)$$

Le modèle IRT à 3 paramètres a donc été sélectionné pour prendre en compte à la fois la difficulté des items mais aussi la probabilité d'obtenir la bonne réponse par une sélection aléatoire.

Les quiz proposés par l'application mobile se prêtent particulièrement à l'application de la théorie IRT car il s'agit de QCM avec un volume de réponses enregistrées conséquent, permettant d'avoir matière à faire converger les algorithmes d'estimation des paramètres. À l'aide du package *mirt* de R et des réponses des apprenants à chaque question, il est donc possible d'obtenir pour chaque question les valeurs a , b et c et d'estimer ensuite le niveau de capacité θ des apprenants ayant répondu aux quiz.

À partir des valeurs des paramètres a , b , c de chaque question, nous cherchons ensuite à connaître le niveau de capacité réel des apprenants sur le quiz dans son intégralité. Pour ce faire, considérant la séquence des réponses apportées aux questions d'un quiz par un apprenant $S = [vrai, faux, \dots, vrai]$ et la formule de l'équation 1 avec les paramètres a , b , c préalablement estimés, nous pouvons calculer la probabilité $P(S|\theta)$ en considérant chaque

question du quiz comme un événement indépendant. En effet, les quiz sont conçus tels que les questions peuvent être abordées dans n'importe quel ordre et la réponse apportée à une question n'influe pas sur une autre question. On a donc la formule 2 :

$$P(S|\theta) = \prod_i P_i(v|\theta) \quad (2)$$

avec S le vecteur des valeurs des réponses aux questions du quiz, i la i -ème question du quiz, v la valeur de la réponse ($v = 0$ si la réponse est fautive, 1 sinon) et P_i la probabilité de répondre correctement à l'*item* selon la formule 1.

En initialisant $\theta = -4$ et en réutilisant la procédure itérative basée sur le maximum de vraisemblance décrite par (2001), nous assignons à l'apprenant la valeur θ qui maximise $P(S|\theta)$. À partir d'une séquence de réponses aux questions d'un quiz donné, nous sommes donc capables de définir un niveau de capacité θ pour l'apprenant sur ce quiz.

3.2. ASSIGNER À L'APPRENANT UNE STRATÉGIE DE RECOMMANDATION

Disposant d'un score de capacité défini par l'IRT pour chaque apprenant ayant terminé un quiz, nous l'avons exploité pour adapter les recommandations.

3.2.1. Répartition des apprenants en trois groupes

Nous souhaitons que le système de recommandations propose des contenus adaptés au niveau de l'apprenant mais la variable θ peut théoriquement prendre toutes les valeurs possibles entre -4 et 4, et il est difficile de concevoir une stratégie de personnalisation aussi fine qui soit pédagogiquement valide en l'absence de prérequis de niveau sur les ressources. En effet, s'il semble assez intuitif de proposer des ressources de révision à des apprenants en difficulté et des ressources d'approfondissement à des apprenants maîtrisant bien le sujet, comment définir la valeur seuil à partir de laquelle un apprenant est estimé en difficulté? Nous avons choisi d'exploiter les seuils existants dans l'application Nomad Education définissant 3 niveaux de difficulté et d'établir une stratégie de recommandation pour chacun de ces 3 niveaux.

Pour définir les seuils séparant les 3 niveaux, pour chaque quiz, toutes les combinaisons possibles de séquences S ont été simulées. Ainsi, pour un quiz de 10 questions, il y a $2^{10} = 1024$ séquences de réponses possibles. Pour chaque séquence, la Formule 2 donne une valeur de θ correspondante. L'intervalle du nombre de bonnes réponses possibles [0; nombre de questions] est séparé en 3. Le seuil entre les niveaux inférieur/intermédiaire est défini comme la moyenne des scores θ obtenus pour les séquences S ayant moins d'1/3 de bonnes réponses. Le seuil entre les niveaux intermédiaire/supérieur est défini comme la moyenne des scores θ obtenus pour les séquences S ayant plus de 2/3 de bonnes réponses. Ceci donne 2 seuils qui sont dépendants des paramètres a, b, c mais indépendants de la distribution des apprenants ayant répondu aux questions du quiz. La Figure 5 montre un exemple de répartition des scores θ pour le quiz *Puissances de 3^{ème}*. Nous voyons que les apprenants ayant un $\theta < -2.873$ sont affectés au groupe de niveau inférieur, et ceux ayant un $\theta > 0.710$ sont affectés au groupe de niveau supérieur. Les apprenants entre ces deux seuils sont affectés au groupe de niveau intermédiaire. Cette figure montre également qu'un apprenant ayant répondu à 7 bonnes réponses peut être dans le niveau intermédiaire ou supérieur selon les questions auxquelles il a répondu. En effet, comme les paramètres IRT a, b, c sont propres à chaque question, une bonne réponse participe plus ou moins à l'évaluation du niveau de capacité. Cette possibilité n'aurait pas été prise en compte en attribuant les niveaux à partir d'un seuil sur les notes brutes obtenues sur les quiz. Dans la suite de ces travaux, le terme *niveau IRT* désignera l'intervalle (inférieur, intermédiaire, supérieur) auquel l'apprenant a été

affecté suite à l'application de cette méthode de séparation des apprenants selon 3 niveaux. La définition de ces seuils et leur impact sera discutée dans la section 8.

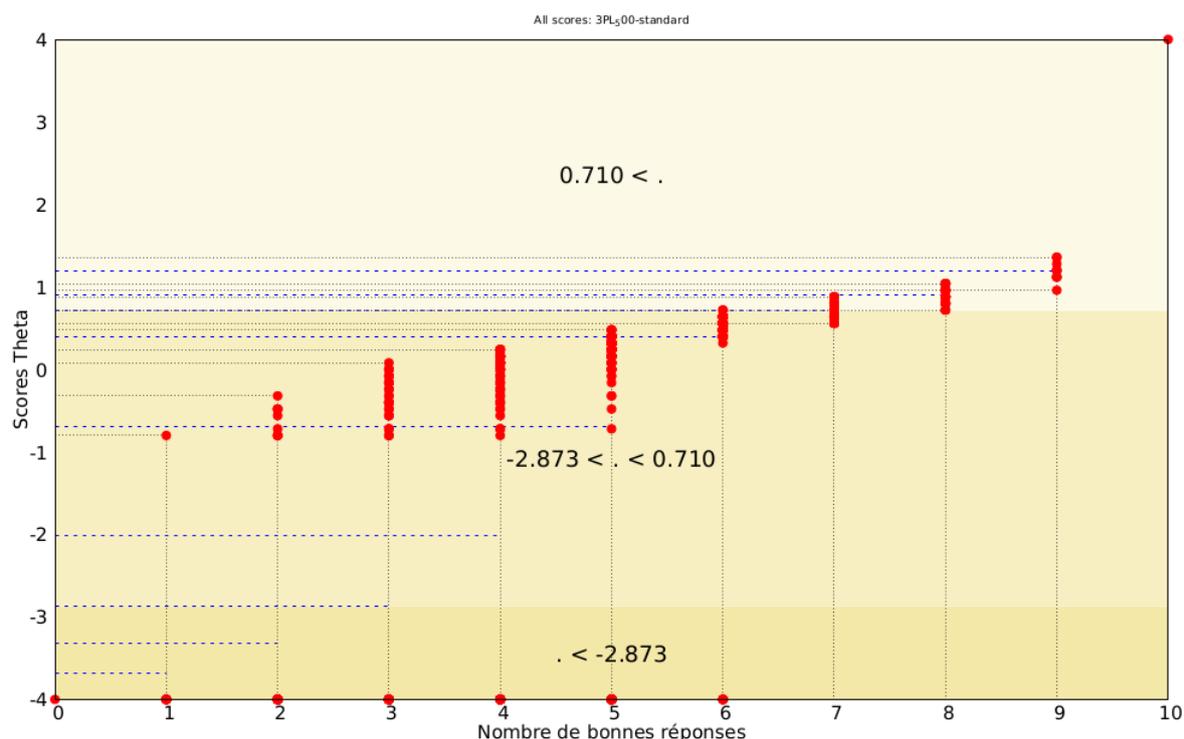


Figure 5 : Répartition des scores θ possibles selon le nombre de réponses correctes pour le quiz *Puissances* 3^{ème} et mise en évidence des seuils IRT. Chaque point rouge correspond à une valeur de θ pour une séquence S. Les pointillés bleus correspondent à la moyenne des θ observés pour les séquences groupées par nombre de bonnes réponses. (Source : <https://www.nomadeducation.fr/>).

3.2.2. Définition de trois stratégies de recommandation

Pour chacun des 3 niveaux de maîtrise possible d'un chapitre (inférieur, intermédiaire, supérieur), nous proposons une stratégie de personnalisation différente en mobilisant le graphe de notions défini en section 2.3. Ces 3 stratégies de recommandation sont pensées selon les hypothèses suivantes :

- Les élèves de niveau IRT inférieur peuvent présenter des difficultés car les prérequis du chapitre étudié sont mal maîtrisés. Ces lacunes peuvent être récentes ou il peut s'agir de difficultés héritées des années scolaires antérieures. Ces élèves seront orientés vers des ressources de niveau scolaire équivalent ou inférieur à son chapitre d'étude actuel, avec des chapitres mobilisant les notions assignées comme prérequis sur le chapitre en cours d'étude. Cette stratégie sera nommée **révision (R)**.
- Les élèves de niveau IRT intermédiaire ne présentent pas de difficultés évidentes, au contraire des élèves de niveau IRT inférieur, mais doivent encore confirmer les savoirs acquis sur le chapitre actuel avec des ressources mobilisant les mêmes notions que le chapitre en cours d'étude, qu'elles soient prérequis ou attendues. Nous leur proposons donc des ressources de niveau équivalent pour poursuivre sur des chapitres liés à leur programme scolaire actuel ou juste un niveau antérieur au chapitre actuel, pour renforcer et remobiliser les savoirs des chapitres vus précédemment. Cette stratégie

sera nommée **continuité** (C).

- Les élèves de niveau IRT supérieur maîtrisent bien les notions attendues du chapitre en cours d'étude. L'enjeu du système de recommandations va donc être d'aller plus loin dans les savoirs en leur faisant découvrir d'autres chapitres qui mobilisent ces mêmes notions attendues, de niveau équivalent ou supérieur au chapitre actuel. Cette stratégie sera nommée **approfondissement** (A).

Afin de définir les chapitres éligibles à la recommandation, les chapitres disponibles dans l'application mobile sont filtrés suivant la stratégie choisie et le graphe de notions. Ce filtrage est illustré par la Figure 6.

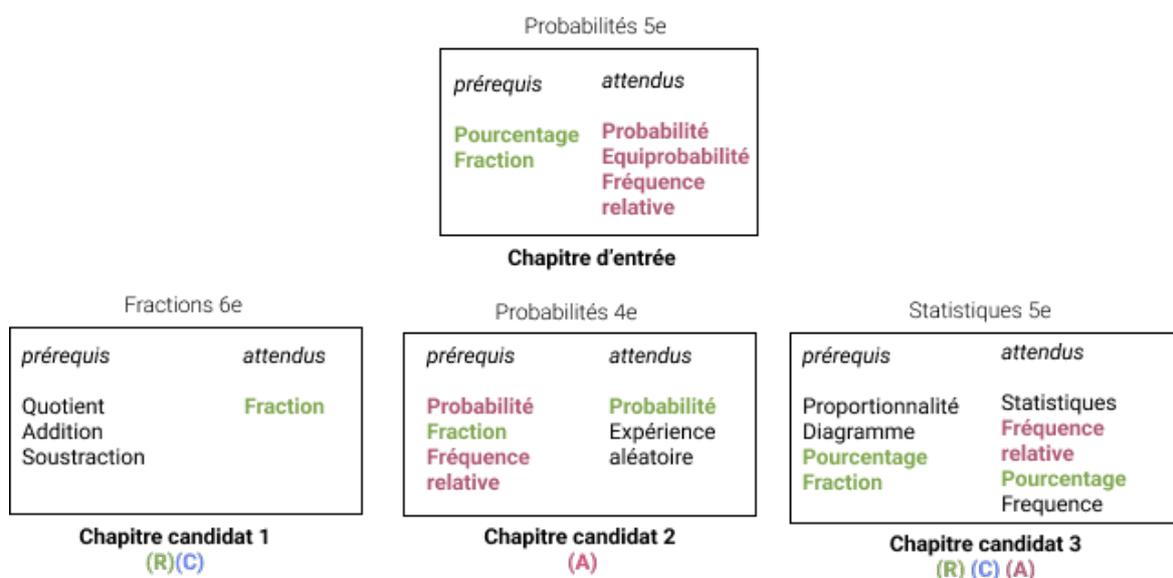


Figure 6 : Exemple de filtrage du graphe de notions selon la stratégie de recommandation sélectionnée. Les notions en commun avec les prérequis du chapitre en cours d'étude sont représentées en vert, celles en commun avec les attendus en rouge. Les chapitres marqués (R),(C) et (A) sont respectivement éligibles pour la stratégie de Révision, Continuité et Approfondissement

Sur cet exemple, nous prenons le cas d'un apprenant qui étudie le chapitre de probabilités de niveau 5^{ème}. Ce chapitre a comme notions prérequisées *Pourcentage* et *Fraction* et comme attendues *Probabilité*, *Equiprobabilité* et *Fréquence relative*.

- Si l'étudiant est affecté au niveau IRT inférieur (**stratégie révision**), tous les chapitres contenant les notions prérequisées du chapitre *Probabilités 5^{ème}* sont conservés, soit *Fraction* et *Pourcentage*. Parmi ces chapitres, seuls ceux de niveau scolaire inférieur ou équivalent au niveau 5^{ème} sont gardés, soit *Fractions 6^{ème}* et *Statistiques 5^{ème}*.
- Si l'étudiant est affecté au niveau IRT intermédiaire (**stratégie continuité**), tous les chapitres liés par les tags prérequis et attendus au chapitre *Probabilités 5^{ème}* sont conservés, et seuls ceux de niveau scolaire équivalent ou juste d'une année inférieure sont gardés, soit ici *Fractions 6^{ème}* et *Statistiques 5^{ème}*.
- Enfin, si l'étudiant est attribué au niveau IRT supérieur (**stratégie approfondissement**), seules les notions attendues du chapitre *Probabilités 5^{ème}* sont prises en compte, soit *Probabilité*, *Equiprobabilité* et *Fréquence relative* et seuls les chapitres de niveau scolaire équivalent ou supérieur sont gardés, soit ici les chapitres *Probabilités 4^{ème}* et *Statistiques 5^{ème}*.

Cet exemple n'est pas exhaustif. Plusieurs chapitres sont ainsi pré-sélectionnés mais seul un nombre réduit est présenté à l'apprenant afin de ne présenter que les plus pertinents mais

également pour tenir compte des contraintes de *design* sur application mobile. Il a donc été nécessaire de définir un mécanisme de classement des ressources éligibles.

3.3. CLASSER LES RESSOURCES SELON 3 COMPOSANTES

L'application mobile Nomad Education est un outil parascolaire et la navigation y est totalement libre. L'objectif du système de recommandations n'était pas de contraindre l'apprenant à étudier des ressources selon un parcours imposé mais de lui laisser cette liberté d'étudier les chapitres de son choix, souvent liés au programme qu'il étudie en parallèle en classe. Le système de recommandations est donc pensé pour pouvoir, après chaque chapitre, recommander des chapitres parmi ceux disponibles dans l'application (appelés chapitres candidats), de niveau adapté et en lien avec le chapitre qui vient d'être étudié par l'apprenant.

Nous avons expliqué comment nous exploitons l'IRT pour définir 3 stratégies de recommandation. Au sein de chaque stratégie, plusieurs chapitres peuvent être sélectionnés, chacun étant plus ou moins pertinent pour l'apprenant. Pour classer ces chapitres retenus, nous nous appuyons sur 3 composantes : une composante de pertinence pédagogique associée à deux composantes historique et nouveauté, afin d'adapter les recommandations à l'utilisation de l'application par les apprenants et maintenir une motivation dans un contexte d'usage non-encadré. Ces trois composantes sont combinées pour calculer un score de recommandation afin d'évaluer la pertinence de chaque chapitre dans un contexte précis.

3.3.1. Prise en compte de la pertinence pédagogique

Le score de pertinence pédagogique s'appuie sur deux indicateurs : d'une part la similarité entre les chapitres, et d'autre part sur l'écart de niveau scolaire entre les chapitres.

Sur l'exemple présenté Figure 6, les chapitres *Fractions 6e* et *Statistiques 5e* sont tous deux éligibles à la recommandation pour les stratégies révision et continuité. Toutefois, le chapitre de *Fraction 6^{ème}* n'a qu'1 seule notion en commun avec le chapitre *Probabilités 5^{ème}* alors que le chapitre *Statistiques 5^{ème}* a 3 notions en commun pour la stratégie révision (tags prérequis du chapitre en cours) et 4 pour la stratégie continuité (tous types de tag confondus). On peut donc penser que le chapitre de *Statistiques 5^{ème}* serait plus en lien avec le chapitre *Probabilités 5^{ème}* que le chapitre de *Fractions 6^{ème}*.

Une possibilité aurait donc été de sélectionner le chapitre qui a le plus de notions en commun avec le chapitre en cours d'étude. Cependant, certaines notions sont plus présentes que d'autres dans le graphe de notions car plus génériques et certains chapitres ont plus de notions apposées que d'autres. Pour prendre en compte ces répartitions inégales des notions, nous utilisons l'indice de similarité *TF-IDF* (*Term-Frequency-Inverse Document Frequency*). Cette métrique issue des modèles vectoriels est couramment utilisée dans les systèmes de recommandations par *content-based filtering* (Lü *et al.*, 2012) ainsi que pour calculer des similarités entre documents ou pages web (Nguyen *et al.*, 2014; Paquette *et al.*, 2015). Elle permet de représenter un document décrit par un vocabulaire sous forme de vecteur. Dans notre cas, notre document est un chapitre, décrit par des notions. Le vecteur $\vec{d}_j = (w_{1,j}, w_{2,j}, \dots, w_{n,j})$ représente pour le chapitre d_j les poids $w_{i,j}$ associés à chaque notion i parmi les notions présentes sur le chapitre en cours d'étude (notions prérequis si stratégie révision, attendues si stratégie approfondissement, toutes notions sinon). Ces poids sont calculés par l'équation 3 :

$$w_{i,j} = TF_{i,j} * \log\left(\frac{N}{df_i}\right) \quad (3)$$

avec $w_{i,j}$ le poids de la notion i dans le chapitre j , $TF_{i,j}$ la fréquence de la notion i dans le chapitre j , df_i le nombre de chapitres pré-sélectionnés contenant la notion i et N le nombre total de chapitres pré-sélectionnés.

Ainsi, plus la notion est générique, *i.e.* présente dans de nombreux chapitres, moins elle a de poids et plus le chapitre a de notions, moins chacune a de poids dans la représentation vectorielle de ce chapitre. Nous calculons ensuite la similarité entre deux chapitres ainsi vectorisés avec la formule de la *similarité cosinus* (équation 4, également utilisée dans la littérature pour calculer une similarité entre documents (Nguyen *et al.*, 2014)).

$$\text{score}_{\text{similarité}_{i,j}} = \text{sim}(\vec{d}_i, \vec{d}_j) = \text{cosine}(\vec{d}_i, \vec{d}_j) = \frac{\vec{d}_i \cdot \vec{d}_j}{|\vec{d}_i| * |\vec{d}_j|} \quad (4)$$

avec \vec{d}_i le vecteur associé au chapitre i défini à partir de la méthode TF-IDF. Pour chaque chapitre et selon chaque stratégie, nous calculons la similarité avec les autres chapitres du graphe de notions pré-sélectionnés par la stratégie.

Outre la similarité par les notions, nous considérons également l'écart de niveau scolaire entre deux chapitres, mesure implicite d'un écart de difficulté. Le programme scolaire français étant structuré en cycles pour les niveaux Primaire/Collège, nous prenons en compte cette spécificité car chaque cycle traduit une cohérence entre les programmes : les niveaux CM1, CM2 et 6^{ème} constituent le cycle 3 dit "cycle de consolidation", et les niveaux 5^{ème}, 4^{ème} et 3^{ème} le cycle 4 dit "cycle des approfondissements" (Education Nationale, 2021). Le lycée forme un cycle indépendant. Ainsi, en se basant sur ces cycles définis par l'Éducation Nationale, nous considérons un écart de niveau plus important entre un chapitre de 6^{ème} et un chapitre de 5^{ème} (inter-cycles) qu'entre un chapitre de 5^{ème} et un chapitre de 4^{ème} (intra-cycle). Cette pénalisation de la distance entre les chapitres est représentée sur l'équation 5 :

$$\text{pénalité}_{\text{distance}} = \frac{c * |L_c - L_i|}{D_{\text{max}}} \quad (5)$$

avec L_c le niveau scolaire du chapitre courant, L_i le niveau scolaire d'un chapitre candidat à la recommandation, D_{max} la distance maximale entre deux niveaux scolaires (7 pour un graphe de notions établi pour les chapitres de 6^{ème} à Terminale) et c un coefficient de pénalité avec $c = 0.25$ pour une distance intra-cycle et $c = 0.75$ pour une distance inter-cycles.

En multipliant ces deux indicateurs (similarité et distance) nous accordons la même importance à la similarité entre deux chapitres qu'à leur écart de niveau scolaire. Le **score de pertinence pédagogique** est ainsi défini par l'équation 6.

$$\mathcal{P} = \text{score}_{\text{pédagogique}} = \text{score}_{\text{similarité}} * (1 - \text{pénalité}_{\text{distance}}) \quad (6)$$

avec $\text{score}_{\text{similarité}} \in [0, 1]$ la similarité entre le chapitre courant et le chapitre candidat déterminé par l'équation 4, et $\text{pénalité}_{\text{distance}} \in [0, 0.75]$ l'écart de niveau scolaire entre les chapitres calculé selon l'équation 5.

3.3.2. Prise en compte de l'historique de l'apprenant

Chaque apprenant ayant une activité propre dans l'application, il est important que le système de recommandations tienne compte de l'historique d'étude de l'apprenant. L'application étant exclusivement parascolaire, nous ne disposons d'aucune information sur les chapitres actuellement étudiés en classe qui permettrait d'orienter la personnalisation. Les seules données accessibles sont les résultats des quiz commencés sur l'application. Afin de prendre en compte l'historique de l'apprenant sur l'application, nous posons plusieurs hypothèses.

Une première hypothèse selon laquelle un chapitre même pédagogiquement pertinent peut sembler inadapté s'il a déjà été réalisé par l'élève. La complétion des chapitres et le moment d'étude de ce chapitre sont donc pris en compte dans la composante historique.

Une seconde hypothèse selon laquelle il peut être bénéfique de proposer à l'élève de retenter un chapitre précédemment échoué si l'acquisition des savoirs sur son chapitre en cours d'étude a permis de débloquent une difficulté. Par conséquent, la note obtenue sur le chapitre pris en compte dans la composante historique.

Une dernière hypothèse selon laquelle on peut raisonnablement supposer que certains acquis s'érodent avec le temps et qu'un rappel de chapitres précédemment étudiés peut être bénéfique. Il apparaît donc important d'inclure une dimension temporelle à cette composante historique afin de ne pas pénaliser définitivement certains chapitres.

Plusieurs travaux en EIAH prennent en compte cet aspect temporel à l'exemple de Choffin *et al.* (2019) qui ont étudié l'effet de la répétition sur l'apprentissage et l'oubli, ou de Jiang *et al.* (2019) qui utilisent un modèle de réseau de neurones récurrents *Long Short Term Memory (LSTM)* pour prendre en compte l'oubli dans leur modèle de prédiction de notes selon l'historique de l'apprenant.

Notre **score historique**, dont le poids va diminuer avec le temps, est donc représenté par l'équation 7. Dans cette équation, la même importance est accordée à la complétion des chapitres candidats qu'à la moyenne des notes obtenues sur les quiz de ces chapitres.

$$\mathcal{H}_t = score_{historique,t} = 1 - (w_t * score_{complétion} * score_{note}) \quad (7)$$

avec $w_t \in [0, 1]$ un facteur qui décroît avec le temps, $score_{complétion} \in [0, 1]$ le ratio de quiz terminés dans le chapitre candidat et $score_{note} \in [0, 1]$ la note moyenne obtenue sur les quiz terminés du chapitre candidat.

Il paraît intuitif de faire diminuer le poids historique des ressources au cours du temps, néanmoins la modélisation de ce facteur temps pose des difficultés. L'étude de travaux sur la rétention mémorielle a montré que la forme de la courbe d'oubli est un problème complexe et débattu en psychologie (Averell et Heathcote, 2011 ; Plaskura, 2019). Les travaux d'Ebbinghaus, fondateurs sur ce sujet, modélisent la courbe de l'oubli suivant une fonction logarithmique (Ebbinghaus, 2013), reprise et adaptée dans plusieurs travaux depuis (Vlach et Sandhofer, 2012). Krauss *et al.* (2016) proposent une amélioration de la courbe d'oubli d'Ebbinghaus avec une modélisation du facteur temps prenant en compte l'effet de répétition, de rétention, le type d'objet mémorisé (les auteurs posent que les vidéos sont davantage mémorisables que le texte) et la difficulté de la ressource à mémoriser. En l'absence de terrain d'expérimentation pour étudier l'impact de tous ces paramètres sur le facteur temps pour l'apprentissage humain dans notre contexte, nous avons choisi de reprendre l'allure de la courbe d'Ebbinghaus pour modéliser le facteur temps mais de prendre des unités de jours et non plus des minutes et d'en adapter les constantes. En effet, Ebbinghaus menait ses expérimentations sur plusieurs minutes avec des exercices de mémorisation de syllabes n'ayant pas de sens. Dans notre cas, nous considérons la mémorisation sur plusieurs jours avec des contenus de chapitres qui doivent être appris et compris. Considérant l'usage non régulier de l'application, le fait qu'il ne s'agisse que d'un outil complémentaire dans leur environnement d'apprentissage et que la quantité de quiz disponibles pour étudier chaque chapitre y est très faible, nous supposons que la rétention y est moindre que celle qu'on obtiendrait en prenant les constantes d'Ebbinghaus (constantes ayant par ailleurs été définies pour faire correspondre la courbe mathématique aux données observées dans le cadre des expérimentations). La formule de la courbe d'Ebbinghaus est donnée par l'équation 8 (Ebbinghaus, 2013). Nous proposons de l'adapter en posant les constantes ($k=1$, $c=5$) afin d'estimer une rétention à 30% au bout de 15 jours et à 12% au bout de 30 jours. Ce choix arbitraire sera discuté dans la section 8.

$$w_t = \frac{100k}{(\log t)^c + k} \quad (8)$$

avec $k=1.84$ et $c=1.25$ dans la formule initiale proposée par Ebbinghaus (2013), avec $k=1$ et $c=5$ dans notre cas.

3.3.3. Faire varier les recommandations

À partir des scores pédagogique et historique, nous sommes en mesure de classer les chapitres éligibles à la recommandation en nous appuyant sur la structure du graphe de notions défini par les professeurs, et en prenant en compte le niveau de l'apprenant *via* l'IRT, ainsi que l'historique de celui-ci dans l'application. Dans notre contexte d'usage volontaire et non-encadré, la motivation de l'apprenant est un facteur déterminant comme le montrent les travaux sur les MOOCs (Kizilcec *et al.*, 2013 ; Reich, 2014), ce pourquoi nous avons ajouté une composante nouveauté afin d'introduire une diversité dans les recommandations et de remédier ainsi à un éventuel effet de lassitude (Castells *et al.*, 2022). Le terme de sérendipité revient dans de nombreux travaux pour prendre en compte cet aspect mais sa définition est débattue : tous s'accordent à définir la sérendipité comme une offre de recommandations pertinente mais inattendue, mais certains incluent dans cette définition d'autres critères comme la diversité, la nouveauté, la pertinence, etc. (Kotkov *et al.*, 2016 ; Ziarani et Ravanmehr, 2021). Nous avons donc choisi le terme de nouveauté et nous pénalisons les chapitres ayant déjà été recommandés par le système à l'apprenant, en pénalisant davantage les chapitres recommandés en première position que ceux recommandés en dernière position. De même que pour la composante historique, un facteur temporel est introduit pour permettre de recommander à nouveau un chapitre ayant déjà été proposé il y a longtemps. Afin de limiter le nombre de paramètres dans le modèle et sans étude préalable sur l'évolution de la rétention mémorielle dans un contexte de navigation sur application mobile, nous avons repris le même facteur temps que celui utilisé pour la composante historique. Le **score de nouveauté** est défini par l'équation 9 :

$$\mathcal{N}_t = \text{score}_{\text{nouveauté}, t} = 1 - [w_t * \frac{1}{n} \sum_{i=1}^n \frac{M - k_i}{M}] \quad (9)$$

avec n le nombre de fois où le chapitre a été recommandé, t le temps depuis la dernière fois où le chapitre a été recommandé, $k_i \in [0, M[$ le rang de la ressource dans le top-M des recommandations proposées (0 étant le meilleur) la i^e fois et $w_t \in [0, 1]$ le facteur temps.

3.3.4. Combinaison des différentes composantes en un score de recommandation

Pour définir notre modèle de recommandation, nous avons posé l'hypothèse que les recommandations doivent être pédagogiquement cohérentes, c'est-à-dire aborder des savoirs communs avec ceux du chapitre actuellement étudiés et être de niveau de difficulté proche. Cette hypothèse s'inspire du concept de Zone Proximale de Développement, et nous a permis de proposer un score de pertinence pédagogique (équation 6). Partant du constat que les notions apposées sur les chapitres dans le graphe de notions étaient plus ou moins génériques, et que chaque chapitre avait un nombre variable de notions, le calcul de similarité par TF-IDF utilisé dans plusieurs systèmes de recommandations de la littérature a été exploité pour définir une similarité pédagogique entre les chapitres. Le système de recommandations se voulant transversal aux niveaux scolaires afin de permettre aux apprenants d'étudier

selon leur besoins et de faire des liens entre les chapitres de toute leur scolarité, des chapitres similaires au sens TF-IDF peuvent se trouver éloignés en terme de niveau scolaire et donc implicitement de niveau de difficulté. Pour prendre en compte le niveau du chapitre recommandé et l'adapter à la stratégie de personnalisation liée au niveau de l'élève, le score pédagogique pénalise les chapitres selon la distance au niveau scolaire du chapitre en cours d'étude. Le calcul de cette pénalisation s'appuie sur le programme de l'Éducation Nationale qui regroupe les niveaux scolaires en cycles.

Nous avons complété notre modèle de recommandation en nous appuyant sur 2 autres hypothèses. La première est que l'historique de l'apprenant doit être intégré au système pour ne pas le reconfronter trop tôt à des échecs et pour ne pas l'ennuyer avec des chapitres qu'il maîtrise déjà ou qu'il vient d'étudier au sein de l'application. Pour ce faire, la formule du score historique (équation 7) pénalise les chapitres récemment étudiés et bien maîtrisés par l'apprenant. En nous appuyant sur plusieurs travaux de la littérature en EIAH et en psychologie et sur le contexte particulier d'utilisation de l'application mobile Nomad Education, nous avons pris parti de faire diminuer le poids de ce score historique avec le temps. La seconde hypothèse est qu'un système de recommandations qui s'intègre dans une application mobile parascolaire doit proposer des contenus différents afin d'une part d'encourager la curiosité de l'apprenant, d'autre part de tester la pertinence d'autres chapitres qui n'apparaissent pas spontanément dans le top des recommandations classées selon les précédents scores. Cette intuition s'appuie sur les travaux concernant la sérendipité et la diversité dans les systèmes de recommandations. Pour y répondre, un score de nouveauté pénalise les chapitres trop récemment recommandés.

Chacune de ces hypothèses n'a pas le même importance dans nos travaux. La combinaison des formules en résultant en un score de recommandation est fait grâce à une moyenne pondérée avec un poids plus important accordé au score pédagogique ($w=0.5$), un poids moindre au score historique ($w=0.4$) et un poids minimal au score de nouveauté ($w=0.1$). Nous donnons la prévalence à la pertinence pédagogique mais l'historique de travail et de réussite de l'apprenant doivent également être considérés, dans une logique pédagogique mais également de cohérence d'utilisation de l'application. Considérant le nombre limité de chapitres à notre disposition, le score de nouveauté a surtout un rôle dans la variation des recommandations proposées, nous lui accordons le poids minimal. Cette moyenne pondérée avec initialisation manuelle des poids, semblable à ce qui a été fait par Zhao *et al.* (2019) sera discutée en section 8. Pour chaque chapitre et chaque apprenant, à un instant t , le **score de recommandation** des chapitres candidats est calculé selon la formule 10.

$$score_{recommandation,t} = 0.5 * \mathcal{P} + 0.4 * \mathcal{H}_t + 0.1 * \mathcal{N}_t \quad (10)$$

avec \mathcal{P} le score pédagogique, \mathcal{H} le score historique et \mathcal{N} le score nouveauté précédemment définis.

3.4. PROCESSUS IMPLÉMENTÉ ET AFFICHAGE DES RECOMMANDATIONS

Le processus implémenté pour le modèle de recommandation décrit est présenté sur la Figure 7. Ce processus prend deux informations en entrée : le graphe de notions (1 sur la Figure 7) et les résultats de l'apprenant sur les quiz du chapitre en cours d'étude. À partir de ces résultats, un score IRT est calculé (2 sur la Figure 7) qui va permettre de choisir une stratégie de personnalisation (3 sur la Figure 7). Selon cette stratégie, un filtrage des chapitres éligibles à la recommandation est effectué (4 sur la Figure 7). Afin de ne sélectionner que les meilleurs chapitres à recommander à l'apprenant, un classement de ces chapitres pré-filtrés (5 sur la Figure 7) est effectué selon 3 critères : la pertinence pédagogique, l'historique de l'apprenant et le renouvellement des recommandations.

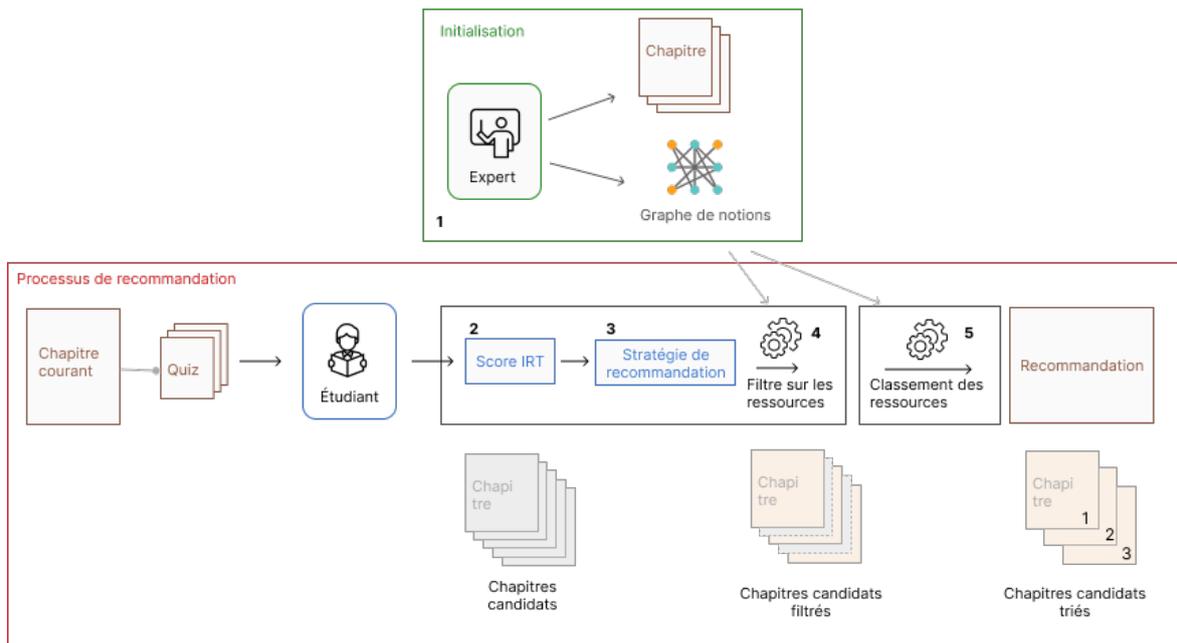


Figure 7 : Processus d'implémentation du système de recommandations

Compte tenu de la taille d'écran disponible sur un smartphone, le nombre de recommandations affichées a été limité à 3. Cela permet de réduire l'impact du *design* sur le choix des recommandations, avec par exemple la dernière recommandation qui pourrait ne pas être affichée sans déroulement explicite de la page d'affichage (*scroll*) par l'utilisateur.

Pour chaque recommandation proposée, l'application affiche le titre du chapitre, le niveau scolaire du chapitre (entre parenthèses) ainsi que les notions de ce chapitre recommandé en commun avec les notions du chapitre en cours (en orange) (voir Figure 8). Cela permet de montrer explicitement à l'apprenant le lien entre le chapitre terminé et les chapitres recommandés, d'une part dans une démarche d'explication des recommandations, d'autre part afin d'inciter les apprenants à repérer eux-mêmes ces liens.

Les recommandations sont proposées à différents endroits dans l'application, chacun de ces emplacements étant pensé pour répondre à un potentiel besoin de l'apprenant.

Lors d'un quiz échoué. Si l'apprenant termine un quiz du chapitre et qu'il n'a obtenu que des réponses fausses sur ce quiz, un bouton sur la page résultats de ce quiz lui propose d'afficher les recommandations de la stratégie révision. En effet, on peut supposer que cet apprenant nécessite une aide immédiate sans attendre les prochains quiz. Ce bouton est optionnel, l'apprenant peut choisir de poursuivre tout de même l'étude du chapitre en cours sans consulter les recommandations (voir écran de gauche de la Figure 1).

Lors d'un parcours interrompu. Les apprenants peuvent choisir de ne pas effectuer les 4 quiz d'un chapitre à la suite, on parle alors de "parcours interrompu". Lorsque l'apprenant demande à quitter le chapitre sans avoir terminé tous les quiz associés, une fenêtre s'ouvre proposant les recommandations de la stratégie calculée selon ses réponses aux quiz terminés du chapitre. Cela permet de répondre à deux possibilités pour expliquer l'abandon du chapitre : l'apprenant interrompt son parcours car le chapitre est trop facile, ou bien il abandonne car ce chapitre est trop difficile. Les recommandations peuvent donc l'accompagner vers des chapitres de niveau adapté (voir écran au centre de la Figure 1).

À la fin du parcours. Lorsque l'apprenant a terminé tous les quiz du chapitre, les recommandations sont proposées pour l'inviter à continuer de travailler sur des chapitres liés, toujours selon la stratégie définie. Cela permet de proposer un accompagnement au-delà de

la note brute sur le chapitre qui sanctionne les résultats obtenus sans proposer de suite (voir écran de droite de la Figure 1).

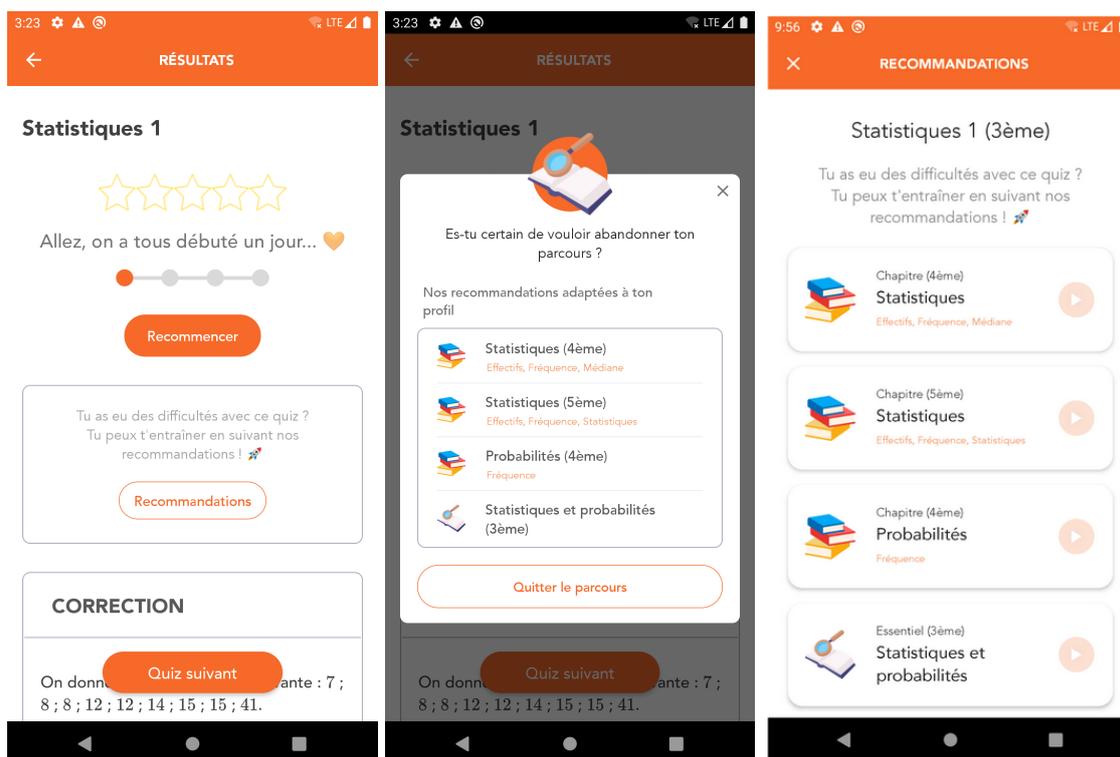


Figure 8 : Interfaces de recommandations pour le chapitre *Statistiques* de niveau 3^{ème} pour la stratégie révision. À gauche, par le point d'entrée quiz échoué, au milieu le cas d'un parcours interrompu et à droite l'interface de fin du parcours. Le niveau scolaire des chapitres recommandés est affiché entre parenthèses. Les notions en commun avec le chapitre actuel sont indiquées en orange. Le contenu "Essentiel" ne fait pas partie du système de recommandations présenté dans le cadre de cet article.

Lorsque l'interface de recommandations occupe tout l'espace écran disponible (cas quiz échoué et fin du parcours), une courte phrase introduit les recommandations au regard de la stratégie assignée. Pour la stratégie révision : "Tu as eu des difficultés avec ce quiz/chapitre ? Tu peux t'entraîner en suivant nos recommandations"; pour la stratégie continuité : "Pas mal ! Découvre nos recommandations pour continuer à progresser"; et pour la stratégie approfondissement : "Bien joué ! Que dirais-tu d'approfondir ces notions?". En raison de l'espace disponible sur l'interface et en concertation avec les équipes *marketing* de l'entreprise sur la manière de communiquer avec le public utilisateur de l'application, cette phrase d'introduction est très courte, elle sous-entend l'objectif de la recommandation mais n'explique pas en détail le principe de recommandation.

4. MÉTHODE D'ÉVALUATION

De nombreux systèmes de recommandations sont utilisés et évalués en classe (Sablayrolles *et al.*, 2022; Venant *et al.*, 2017; Yanes *et al.*, 2020), ou avec des étudiants simulés virtuellement pour disposer d'une quantité de données suffisante pour entraîner les modèles d'apprentissage (Clément *et al.*, 2015), 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.* (2015). Cette difficulté est accentuée dans les contextes d'apprentissage auto-régulés, puisque les apprenants peuvent quitter à tout moment la plateforme d'apprentissage.

Notre modèle de recommandation a été évalué selon plusieurs critères et modalités d'évaluation. Tout d'abord, **des professeurs** ont été invités à évaluer la **pertinence pédagogique des recommandations** au regard des stratégies IRT et du score pédagogique, par l'intermédiaire d'un questionnaire d'évaluation pour la discipline Mathématiques (Badier *et al.*, 2023). Les principaux résultats de cette analyse indiquent que les recommandations proposées sont majoritairement pédagogiquement cohérentes mais les professeurs émettent des réserves sur la pertinence des recommandations proposées aux élèves de niveau supérieur, notamment concernant le rôle d'une application mobile parascolaire pour introduire de nouveaux chapitres aux apprenants. Plusieurs réponses apportées par les professeurs dépassent également le contexte de l'étude, la remarque principale étant "il faut plus d'exercices", ce qui est l'une des contraintes majeures de l'application Nomad Education.

Ensuite, **des apprenants** utilisateurs de l'application ont pu faire un retour direct sur le système de recommandations au moyen d'un questionnaire en ligne. Dans ce questionnaire, les apprenants ont été invités à évaluer le système au regard des **critères d'utilité** (adaptation aux souhaits d'étude ponctuels), de **pertinence** (combinaison des 3 indicateurs), de **variabilité** (indicateur nouveauté), et d'**influence sur le temps passé à étudier** (particularité du contexte de *micro-learning*) (2023). Les réponses à ce questionnaire indiquent que les recommandations sont bien reçues par les apprenants et que leur critère de décision principal pour suivre une recommandation demeure le niveau scolaire des contenus recommandés, ce qui rejoint les remarques faites par les professeurs. On peut néanmoins regretter que le nombre de participants ayant répondu à ce formulaire soit relativement faible au regard du nombre d'apprenants ayant été sollicités pour répondre à cette étude (tous ceux qui avaient accès au système de recommandations), ce qui limite la validité généralisable des résultats obtenus.

Pour valider notre modèle au regard de la spécificité de l'utilisation en contexte de *micro-learning* parascolaire non encadré, les apprenants ont eu la possibilité de tester le système de recommandations en **environnement non contrôlé**. En effet, nous souhaitons bénéficier du contexte industriel qui nous permet de mettre le système directement aux mains des apprenants pour étudier ses usages. Le moteur de recommandations a été implémenté pour la discipline Mathématiques, pour les niveaux scolaires de la 6^e à la Terminale générale. À partir de septembre 2022, le système a été ouvert aux nouveaux utilisateurs de l'application, pour une durée de 1 mois à compter de leur date d'inscription. Nous disposons de cette fenêtre d'observation de 1 mois pendant laquelle les apprenants peuvent accéder librement à cette fonctionnalité qui est ensuite intégrée à la partie payante de l'application mobile.

Pour analyser l'utilisation faite par les apprenants des recommandations proposées, nous avons mis en place un **mécanisme de collecte des traces d'apprentissage** spécifique appelé *AppEvent*. Semblable au système xAPI utilisé dans plusieurs travaux (Branthôme, 2022; Nikolayeva *et al.*, 2018), chaque action de l'utilisateur est identifiée et stockée en base. Nous pouvons ainsi identifier les actions *Vue de l'interface de recommandations* (*AppEvent Recommendation Viewed*), *Clic sur une recommandation de l'interface* (*AppEvent Recommendation Choose*), *Quitter l'interface de recommandations* (*AppEvent Recommendation Left*); 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 corres-

pondante, heure de la consultation 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. Plusieurs analyses ont été conduites pour évaluer ce système de recommandations. Nous allons en présenter trois permettant de répondre aux questions de recherche posées en introduction.

5. FACTEURS INFLUENÇANT LE SUIVI DES RECOMMANDATIONS

Le système de recommandations ayant été implémenté dans une application mobile, nous posons une première question : **Les recommandations proposées sont-elles suivies par les apprenants ?** Les analyses publiées dans Badier *et al.* (2023) ont montré que les recommandations étaient effectivement suivies par des apprenants de différents niveaux scolaires et bien perçues par les utilisateurs de l'application mobile interrogés au regard des critères d'utilité, de pertinence pédagogique et de variabilité.

Nous avons poursuivi l'analyse à partir des observations des données d'utilisation, pour répondre à la question suivante : **QR1- Quels facteurs influencent le suivi des recommandations par les apprenants dans un contexte parascolaire ?**

Nous avons étudié les *AppEvents* collectés entre septembre 2022 et janvier 2023, qui concernent 8 260 apprenants ayant été confrontés au moins une fois à une recommandation, ce qui représente 27 547 traces de recommandations vues, dont 3 927 recommandations suivies par 2 392 apprenants.

Plusieurs facteurs ayant potentiellement une influence sur la décision de suivre ou non une recommandation ont été analysés : la source de la recommandation (quiz échoué, parcours interrompu, fin de parcours), la stratégie (révision, continuité, approfondissement), et le niveau scolaire des chapitres recommandés.

Le premier facteur étudié est la **source de la recommandation**. Pour chaque apprenant les *AppEvents* de type *RecommendationChoose* et de type *RecommendationLeft* sont collectés puis un test statistique du χ^2 est effectué pour déterminer s'il y a statistiquement plus de suivi ou de rejet selon la source de la recommandation. Les résultats de ce test sont présentés dans le Tableau 2 (la colonne *Taux de suivi* a été ajoutée pour faciliter l'interprétation des résultats).

Tableau 2 : Influence de la source de la recommandation sur la décision de suivi (test du Chi-2 de Pearson. $\chi^2 = 1943.5$, $df=2$, $p_{value} < 2.2e - 16$)

Point d'entrée	Recommandation <i>effectifs observés (résidus)</i>			Taux de suivi
	Choisie	Rejetée		
Quiz échoué	661 (+29.9)	872 (-12.2)		43.1%
Parcours interrompu	1188 (-22.4)	14620 (+9.15)		7.5%
Fin du parcours	2078 (+16.3)	8128 (6.66)		20.4%

Ces résultats indiquent que les recommandations sont plus suivies lorsqu'elles apparaissent à la suite d'un quiz échoué (résidus +29.9) ou de parcours terminé (résidus +16.3).

Les recommandations issues de parcours interrompus sont plus rejetées que les autres (résidus -22.4). Il faut toutefois souligner que dans le cas des quiz échoués, l'accès aux recommandations se fait par une démarche explicite de clic sur le bouton "Recommandations" alors que ce n'est pas le cas pour les autres sources où l'affichage de la page est automatique. Cela peut expliquer le meilleur taux de suivi dans cette situation.

Le second facteur étudié est l'influence de la **stratégie de recommandation** appliquée sur le choix de suivre ou non une recommandation. Les résultats du test du χ^2 sont présentés dans le Tableau 3. Nous constatons, toujours grâce aux résidus du test, que les recommandations associées à la stratégie de révision ont tendance à être plus suivies (résidus +9.83) que les stratégies de continuité ou d'approfondissement.

Tableau 3 : Influence de la stratégie de recommandation sur la décision de suivi (test du Chi-2 de Pearson, $\chi^2 = 156.33$, $df=2$, $p_{value} < 2.2e - 16$)

Stratégie	Recommandation <i>effectifs observés (résidus)</i>			Taux de suivi
	Choisie	Rejetée		
Révision	1201 (+9.83)	5149 (-4.0)		18.9%
Continuité	1288 5.86)	(- 9349 (+2.39)		12.1%
Approfondissement	1438 1.74)	(- 9122 (+0.71)		13.6%

Nous nous intéressons ensuite au facteur du **niveau scolaire des chapitres recommandés** puisque ce critère a été identifié comme important par les apprenants ayant répondu à l'étude (2023). Nous collectons les *AppEvent* de type *RecommendationView* et *RecommendationChoose*, et comparons pour chaque stratégie les niveaux scolaires des recommandations vues et des recommandations choisies aux niveaux scolaires des chapitres en cours d'étude. Les résultats sont présentés dans le Tableau 4.

Nous observons que dans chacune des 3 stratégies, les chapitres de niveau équivalent sont proportionnellement plus suivis que les chapitres différents du niveau scolaire actuel. En effet, pour la stratégie révision même si le modèle recommande davantage de chapitres de niveau inférieur que de niveau équivalent au chapitre actuel (2 480 contre 968), 44.9% des chapitres de niveau équivalent sont adoptés contre seulement 30.8% des chapitres de niveau inférieur. Cette même tendance s'observe pour les autres stratégies.

Ces analyses de données confirment donc les retours des professeurs et des apprenants : les recommandations sont davantage suivies si elles interviennent pour proposer une remédiation (source quiz échoué et stratégie de révision), et les élèves suivent davantage une recommandation correspondant à leur niveau scolaire actuel qu'une recommandation de niveau différent.

Tableau 4 : Niveaux scolaires des chapitres choisis parmi les chapitres disponibles à la recommandation pour les recommandations suivies. L = Niveau du chapitre actuel. (Les données manquantes sont dues à des problèmes techniques de collecte des *AppEvents*.)

Stratégie	Chapitres		
	Choisis	Recommandés	Choisis/Recommandés
Révision			
Niveau inférieur ($< L$)	765	2480	30.8%
Niveau équivalent (L)	435	968	44.9%
<i>Données manquantes</i>	1		
Continuité			
Niveau inférieur ($L - 1$)	617	2066	29.9%
Niveau équivalent (L)	669	1595	41.9%
<i>Données manquantes</i>	2		
Approfondissement			
Niveau équivalent (L)	683	1752	39.0%
Niveau supérieur ($> L$)	755	2459	30.7%

6. EFFET DES RECOMMANDATIONS SUR L'EXPÉRIENCE D'APPRENTISSAGE

Après avoir identifié certains critères de sélection des recommandations, nous nous intéressons à la deuxième question de recherche : **QR2 - Quels usages des recommandations pouvons-nous observer dans une application parascolaire ?**

Dans un premier temps nous étudions les **activités réalisées au sein des chapitres recommandés suivis**. Pour cela, nous reprenons les 3 927 traces des 2 392 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 5.

Tableau 5 : 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=3927
Aucune activité	1302 (33.2%)
Au moins 1 cours consulté	1073 (27.3%)
Au moins 1 quiz commencé	2144 (54.6%)
Au moins 1 quiz commencé et terminé	1392 (35.4%)

Au sein de chaque chapitre, l'apprenant peut étudier 2 ressources différentes : les quiz et les cours. Nous constatons que 33.2% des recommandations suivies n'ont débouché sur aucune activité dans le chapitre suivi : soit parce que la recommandation ne correspondait pas aux attentes des apprenants, soit 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 exploratoire*. Dans la majorité des cas, la recommandation suivie a mené à une réelle activité : les apprenants ont commencé au moins un quiz (54.6% des cas) ou consulté au moins un cours (27.3% des cas) (*suivi partiel*) et dans 35.4% des situations analysées, 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'**effet du suivi des recommandations dans l'expérience d'apprentissage**. Dans notre contexte d'utilisation volontaire et parascolaire, l'objectif n'est pas de mesurer un gain d'apprentissage mais plutôt des évolutions de comportements de travail avec l'application. Pour qualifier l'expérience d'apprentissage, nous définissons plusieurs indicateurs calculés à partir des *AppEvents* : *nombre de sessions de travail, durée des sessions de travail, nombre de quiz commencés, nombre de quiz terminés, nombre de cours consultés et temps moyen de lecture des cours*.

Nous sélectionnons les données de sessions complètes de 1 126 apprenants ayant commencé au moins un quiz de Mathématiques dans la session, afin de comparer l'expérience d'apprentissage entre des sessions avec suivi et sans suivi des recommandations, quelle que soit la source ou la stratégie de cette recommandation. Sont donc exclus de cette étude les apprenants pour lesquels nous n'avons que des sessions de travail avec suivi ou que des sessions de travail 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 6.

Tableau 6 : 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 de Mathématiques commencés par session	3.97 ***	2.28	<2.2e-16
Nombre moyen de quiz de Mathématiques terminés par session	3.27 ***	1.85	<2.2e-16
Nombre moyen de cours de Mathématiques consultés par session	3.01 ***	1.67	<2.2e-16
Temps moyen de lecture des cours de Mathématiques 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 *micro-learning*. Les apprenants 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.

7. ANALYSES DES MOTIFS DE NON-SUIVI DES RECOMMANDATIONS

Afin de comprendre comment nous pourrions améliorer le taux de suivi des recommandations, nous nous intéressons à la troisième question de recherche : **QR3 - Peut-on mettre en évidence des leviers d'amélioration du modèle de recommandation à partir des analyses de traces issues de la navigation libre des apprenants sur l'application ?**

Nous analysons donc plus en détail les situations où l'apprenant ne suit aucune recommandation sur l'interface présentée pour naviguer librement vers une autre ressource. Nous identifions l'**activité qui suit une recommandation non-suivie** pour comprendre dans quelle mesure le système de recommandations aurait pu proposer cette activité.

Nous collectons les données de 5 281 apprenants ayant rejeté au moins une recommandation de Mathématiques entre septembre 2022 et janvier 2023, ce qui représente 13 864 traces d'utilisation. Plusieurs actions utilisateurs dans l'application ont été observées :

- (A1) : l'apprenant quitte l'application ;
- (A2) : l'apprenant continue à travailler mais sur une autre discipline ;
- (A3) : l'apprenant continue à travailler la même discipline mais avec un autre chapitre que celui qui est recommandé ;
- (A4) : l'apprenant continue à travailler sur le chapitre en cours.

La Figure 9 représente la répartition des activités enregistrées consécutives au non-suivi des recommandations.

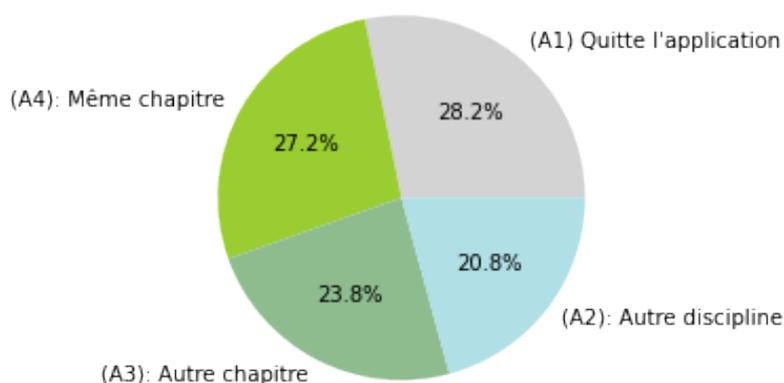


Figure 9 : Activités réalisées après un non-suivi des recommandations, d'après l'analyse des *AppEvents* (N=13 864)

Dans 28.2% des 13 864 cas analysés, les apprenants ne suivent pas les recommandations et quittent l'application (A1). On peut supposer que, quelle qu'ait été la recommandation, les apprenants n'ont pas davantage de temps à consacrer à l'application (contexte d'apprentissage en *micro-learning*). Le même raisonnement peut être appliqué aux 20.8% des situations où les apprenants délaissent les recommandations pour aller travailler une autre discipline

(A2) : le système ne proposant que des recommandations intra-disciplinaires, il ne peut répondre à ce besoin.

Dans les 51% des situations restantes, les apprenants continuent à travailler sur la discipline en cours. Dans 27.2% des cas, ils décident de continuer à travailler sur le même chapitre (A4), ce qui sort également du cadre d'application du système de recommandations puisque le système ne recommande jamais le chapitre courant.

Nous nous intéressons donc plus précisément aux 23.8% des apprenants qui continuent à travailler sur la même discipline mais avec un autre chapitre (A3). Dans cette situation, l'apprenant accepte de changer de chapitre pour continuer à travailler sur la discipline mais refuse les recommandations : on peut supposer qu'il est possible de proposer des recommandations pertinentes à ces apprenants mais que les recommandations actuelles n'ont pas répondu à leur besoin.

Nous cherchons à savoir si cette répartition diffère selon la source (recommandation vue après un quiz échoué, un parcours interrompu ou un parcours terminé) ou selon la stratégie recommandée.

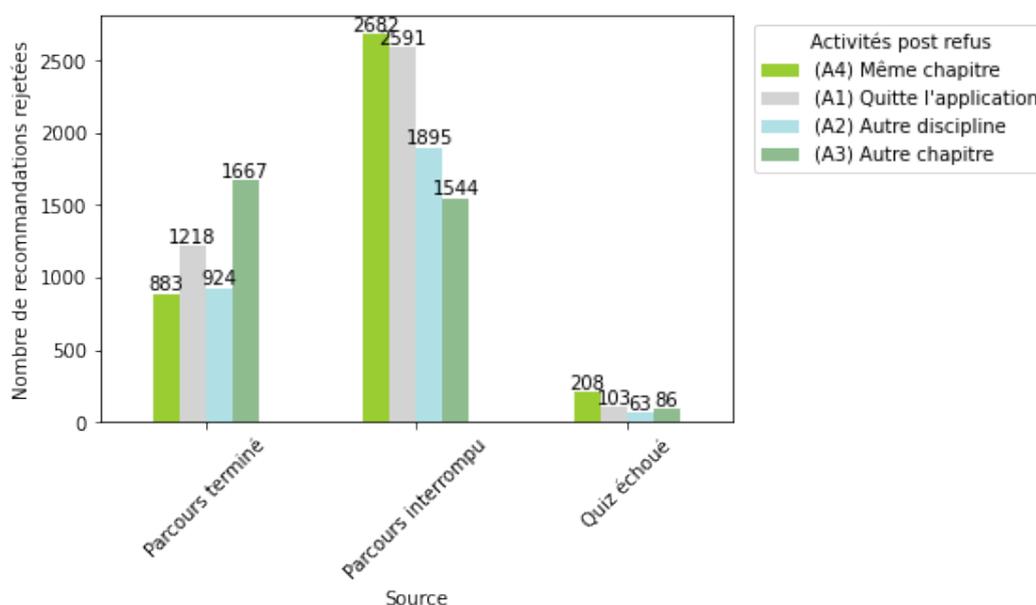


Figure 10 : Répartition des activités réalisées après un non-suivi des recommandations, selon la source de la recommandation vue

Sur la Figure 10, nous nous intéressons à la répartition des activités consécutives au non-suivi des recommandations, selon la source de la recommandation vue. On constate que lorsque les recommandations sont affichées après un parcours terminé, dans la plupart des cas, les élèves continuent à travailler sur un autre chapitre de la même discipline (A3) : dans ces situations, les recommandations sont donc potentiellement améliorables puisque les élèves n'abandonnent pas l'étude de la discipline.

Lorsque les recommandations sont affichées après un parcours interrompu, dans la majorité des situations, soit l'apprenant se ravise et revient sur son chapitre en cours d'étude (A4), soit il quitte l'application (A1), ou encore travaille une autre discipline (A2). On ne peut donc pas affirmer dans ces situations précises, que des recommandations différentes permettraient de maintenir l'élève dans l'étude de cette discipline avec d'autres chapitres. Toutefois, pour 18% (1 544 cas) des situations, l'élève poursuit l'étude de la discipline avec d'autres chapitres (A3), c'est donc le contenu des recommandations qui est inadapté aux

souhait de l'élève.

Lorsque les recommandations sont affichées sur demande de l'élève suite à un quiz échoué, dans 45% des situations, l'élève se ravise et poursuit son parcours au sein du même chapitre (A4).

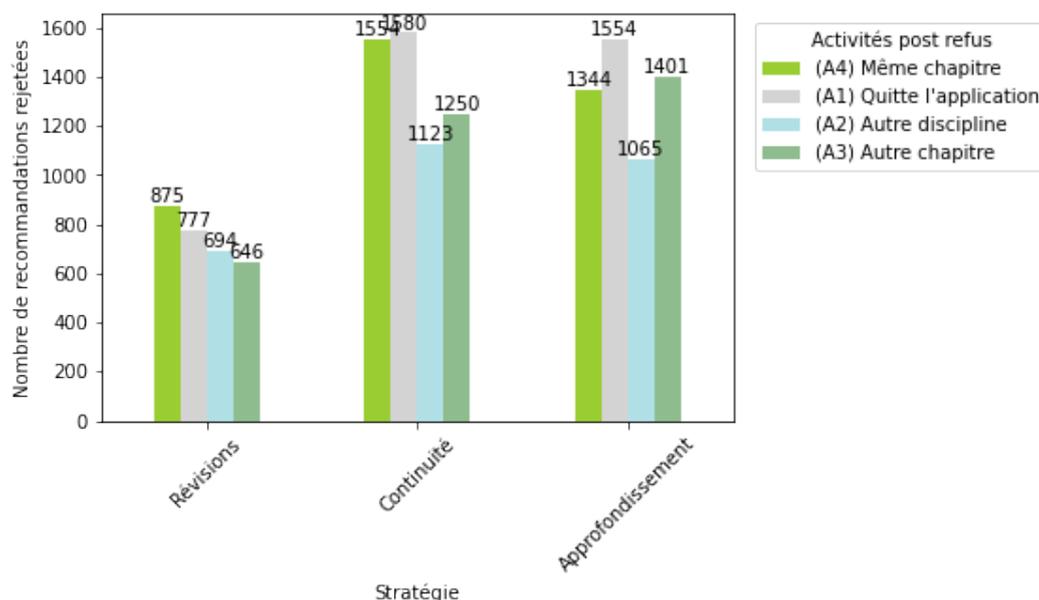


Figure 11 : Répartition des activités réalisées après non-suivi des recommandations selon la stratégie de la recommandation vue

Sur la Figure 11, nous étudions la répartition des activités consécutives au non-suivi d'une recommandation selon la stratégie recommandée. Quelle que soit la stratégie appliquée, les actions A1 et A4 sont les principales actions relevées. Quelle que soit la stratégie de recommandation appliquée suite aux résultats de l'apprenant, certains refus se poursuivent par l'étude d'un autre chapitre de la même discipline (A3), ce sont donc des situations où le système de recommandations est améliorable. D'après les répartitions des activités consécutives au non-suivi des recommandations, relativement similaires entre les stratégies d'après la Figure 11, l'activité qui suit le non-suivi d'une recommandation ne semble pas directement liée à la stratégie assignée.

Après avoir observé l'activité des apprenants en globalité, nous décidons d'étudier plus en détails l'action A3 en essayant de comprendre la raison pour laquelle le chapitre sur lequel les apprenants ont choisi de travailler n'a pas été recommandé par le système. Les résultats sont présentés sur la Figure 12.

Dans 60.8% des cas, le chapitre que l'élève a choisi n'est pas lié au chapitre en cours d'étude d'après le graphe de notions, quel que soit le type de tag considéré (prérequis/attendu). Le graphe de notions étant une donnée d'entrée de notre modèle de recommandations, établi par des professeurs, ce constat a de nombreuses implications qui seront discutées dans la section 8.

Nous nous intéressons aux 39.2% des cas restants, où c'est le système de recommandations qui est directement concerné. Dans 5.2% des situations, l'apprenant a choisi d'étudier un chapitre qui est bien relié au chapitre courant dans le graphe de notions, mais qui n'a pas été filtré par la stratégie assignée par l'IRT. Dans 25.5% des situations, la recommandation choisie par l'apprenant est bien dans la liste des recommandations possibles, mais en dehors du top-3 pédagogique pour la stratégie à laquelle il a été assigné. La combinaison avec les

scores *historique* et *nouveauté* n'a donc pas suffi à faire remonter le score de recommandation du chapitre choisi par l'élève (seuls les 3 premiers chapitres sont recommandés et affichés à l'écran de l'application). On identifie donc que dans 30.7% des situations où les apprenants déclinent les recommandations pour étudier d'autres chapitres de la discipline que ceux proposés, c'est la stratégie de recommandation combinée au score pédagogique qui serait à modifier pour essayer d'améliorer le taux de suivi des recommandations. Dans 8.5% des situations, le chapitre choisi par l'apprenant était bien dans le top 3 des recommandations, si l'on regarde uniquement la composante *pédagogique* du score pour la stratégie assignée, mais elle n'a pas été recommandée à l'apprenant car les modules *historique* et *nouveauté* ont pénalisé cette recommandation par rapport à d'autres chapitres au score de pertinence pédagogique proche.

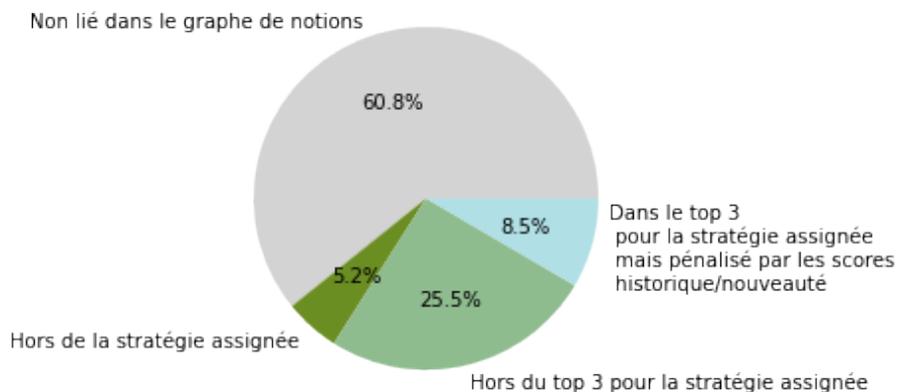


Figure 12 : Raisons pour lesquelles le système n'a pas recommandé le chapitre choisi par l'apprenant dans le cas des recommandations non-suivies avec activité A3

Cette étude des activités consécutives au non-suivi des recommandations ouvre donc la possibilité de concevoir de modèles alternatifs au système de recommandations proposé en partant des analyses de traces d'usages et en posant l'hypothèse que les choix des apprenants identifiés sont des choix pertinents.

8. DISCUSSION

La mise en place du moteur de recommandation dans une application mobile parascolaire et l'analyse de son utilisation par les apprenants a nécessité de faire des choix. Nous revenons à présent sur certains d'entre eux pour les discuter.

8.1. UTILISATION DE L'IRT POUR DÉFINIR DES STRATÉGIES DE RECOMMANDATION

La méthode de l'Item Response Theory occupe une place importante dans l'approche proposée, puisque c'est à partir du niveau de capacité θ , calculé par le modèle IRT à 3 paramètres sur les quiz résolus par l'apprenant, que se fondent les stratégies de personnalisation. Si classiquement dans la littérature scientifique, l'IRT est utilisée pour concevoir des tests adaptatifs, nous en avons adapté l'usage pour associer à un score θ un besoin de recommandation. Le modèle IRT utilisé est un modèle unidimensionnel : le θ calculé à partir des réponses aux quiz d'un chapitre donné s'interprète comme le niveau de capacité de l'apprenant sur le chapitre donné dans son ensemble. Un chapitre mobilisant plusieurs composantes

de connaissances (les notions du graphe de notions défini par les professeurs), utiliser un modèle d'IRT multi-dimensionnel permettrait de définir un vecteur de scores $(\theta_1, \theta_2, \dots, \theta_k)$ associant un niveau de maîtrise différencié à chaque notion mobilisée dans le chapitre étudié. L'utilisation d'un modèle multi-dimensionnel permettrait donc de fournir un diagnostic plus précis, différencié par notions. Néanmoins, une approche multi-dimensionnelle ne peut se faire qu'avec l'appui d'un ou plusieurs didacticiens pour confirmer les notions apposées sur les chapitres. Dans un contexte industriel et avec l'objectif de proposer un modèle de recommandation indépendant de la discipline, la mobilisation de tant de ressources humaines est difficilement envisageable. De plus, la dimensionnalité des tests IRT est un sujet complexe, discuté par Juhel (1999) (et également questionné dans le cadre des tests PISA par Elodie et Monseur (2018)). Par ailleurs, on peut raisonnablement penser qu'avec le faible nombre de ressources et le graphe de notions appliqué uniquement au niveau des chapitres (et non pas descendu jusqu'aux questions), nous n'aurions pas eu de stratégie de personnalisation avec une granularité assez fine à faire correspondre à un diagnostic plus fin des compétences.

De plus, les seuils entre les différents niveaux IRT ont été définis arbitrairement. Les analyses ont montré d'une part que certaines stratégies étaient davantage plébiscitées par les apprenants et les enseignants, et d'autre part, que certains apprenants ne suivaient pas les recommandations pour aller explorer librement des chapitres en dehors de la stratégie assignée par le score IRT. Il pourrait donc être bénéfique d'ouvrir le système aux apprenants pour leur permettre d'accéder aux recommandations des autres stratégies que celle à laquelle ils ont été assignés. Nous pourrions également envisager une sélection des chapitres à recommander non pas selon un critère strict dépendant d'un seuil séparant les niveaux, mais plutôt selon une probabilité d'appartenance à l'une ou l'autre des stratégies d'après le score IRT.

8.2. GRAPHE DE NOTIONS

Le graphe de notions a été constitué grâce à l'annotation des chapitres par un unique annotateur par chapitre, le professeur rédacteur du chapitre. Il aurait été bénéfique de renforcer la robustesse de ce graphe en proposant à plusieurs professeurs d'annoter les mêmes ressources, afin de vérifier la cohérence des tags. En outre, l'analyse des motifs de non-suivi des recommandations a mis en évidence certains liens entre chapitres faits par les apprenants et identifiées dans les traces de leur navigation libre. Certains de ces liens n'étaient pas présents dans le graphe de notion initial, d'autres possiblement insuffisamment pondérés car le chapitre choisi n'avait pas un score pédagogique suffisamment haut pour être recommandé par le modèle de recommandation (voir Figure 12). L'élaboration des graphes de connaissances est un défi important dans les travaux autour de l'Adaptive Learning (Allègre *et al.*, 2021 ; Hubert *et al.*, 2022) et cet aspect mériterait une étude approfondie.

8.3. FACTEUR TEMPS

Un facteur temps a été introduit à deux reprises, dans les scores *historique* (voir équation 7) et *nouveauté* (voir équation 9) et peut être discuté. La valeur de ce facteur temps étant la même dans les deux scores, afin de limiter le nombre de paramètres du modèle à ajuster, cela présuppose que l'oubli des savoirs liés à un chapitre évolue de la même façon que l'oubli des recommandations déjà proposées sur l'application. Ce postulat très fort est en réalité peu probable, néanmoins nous ne disposons pas de données expérimentales permettant de mesurer précisément cette évolution temporelle de la mémorisation dans chacune des situations, ni de les différencier. De même, on peut raisonnablement penser que le taux de mémorisation est propre à chaque élève et propre à chaque contenu, à l'exemple de Krauss *et al.* (2016) qui posent le type et la difficulté des ressources comme des paramètres de la courbe de rétention.

Nous nous intéressons ici préférentiellement à l'allure de la courbe plutôt qu'aux valeurs de rétention, ce pourquoi nous avons choisi de reprendre une équation de courbe existante dans la littérature plutôt qu'une approximation linéaire qui aurait été plus simple mais tout aussi arbitraire et moins justifiable.

8.4. SCORE DE CLASSEMENT DES RECOMMANDATIONS

Le score de recommandation des ressources proposé s'appuie sur 3 composantes : *pédagogique* (\mathcal{P} défini par l'équation 6), *historique* (\mathcal{H}_t défini par l'équation 7) et *nouveauté* (\mathcal{N}_t défini par l'équation 9). Si l'indicateur *pédagogique* est relatif aux ressources et non pas directement aux apprenants, les deux derniers indicateurs dépendent de l'utilisation de l'application par chaque apprenant. L'objectif de ces scores étant de classer les ressources éligibles à la recommandation les unes par rapport aux autres, la valeur chiffrée du score de recommandation n'est pas utilisée en tant que valeur absolue mais relativement aux valeurs des autres ressources, ce qui limite l'impact des valeurs des coefficients sur la recommandation finale. Ces 3 composantes sont combinées dans une formule pondérée pour calculer un score global de recommandation (voir équation 10) dont les poids ont été définis arbitrairement pour laisser une place plus importante au score *pédagogique*, moindre pour le score *historique* et minimale pour le score de *nouveauté*. Nous pourrions envisager une paramétrisation différentes de ces poids, ou encore de faire des *ablation studies*, c'est-à-dire proposer pour certains utilisateurs des modèles de recommandations sans prise en compte d'un ou plusieurs des scores *pédagogique*, *historique* ou *nouveauté*, ceci afin de quantifier l'apport des différents scores dans la recommandation proposée. Nous pourrions également utiliser une fonction *softmax* afin de permettre un renouvellement des recommandations sans passer par la définition d'un indicateur \mathcal{N}_t . Cet indicateur permet néanmoins de prendre en compte le rang de la recommandation précédemment affichée.

8.5. IMPACT DE LA PÉRIODE D'ANALYSE

L'application étant destinée à un usage volontaire et non encadré, les évaluations ont été menées dans ce même contexte, à partir des traces d'utilisation collectées. Le système de recommandations a été mis aux mains de nouveaux utilisateurs de l'application, pour éviter le biais d'attrait pour la découverte d'une nouvelle fonctionnalité chez les utilisateurs existants. Néanmoins, cette option pose également des inconvénients car de nombreux utilisateurs installent l'application, l'explorent puis ne s'en servent plus du tout. Notre contexte d'utilisation rend donc très difficile la personnalisation, puisque le suivi des apprenants est très limité et imprévisible. De plus, nous savons que la période de l'année scolaire a une influence sur les motivations des apprenants (on constate un usage démultiplié de l'application les semaines précédant les examens nationaux). On pourrait imaginer paramétrer le système de recommandations selon la période de l'année, par exemple proposer des révisions des années antérieures en début d'année scolaire et restreindre les recommandations aux contenus de l'année courante au moment des examens nationaux. Si de nombreux éléments prêtent à penser que ce fonctionnement répondrait davantage aux attentes des utilisateurs de l'application mobile, cette approche ne permettrait pas de répondre au besoin de personnalisation individualisée et ne s'inscrit pas dans l'objectif initial du système, qui a été pensé pour proposer des recommandations transversales aux niveaux scolaires, afin de répondre avant tout à un besoin pédagogique identifié chez un élève particulier.

9. CONCLUSION ET PERSPECTIVES

Dans cette étude, nous avons présenté un modèle de recommandation de ressources pédagogique intégré au sein d'une application mobile parascolaire ainsi que ses analyses d'usages. Le modèle de recommandation présenté s'appuie sur trois stratégies de recommandation : *révision*, *continuité* ou *approfondissement*, et sur un score de recommandation qui classe les chapitres sélectionnés selon la stratégie, pour ne recommander que les meilleurs. Ce score de recommandation s'appuie sur un indicateur *pédagogique* qui vérifie la pertinence du chapitre par rapport au chapitre courant (en termes d'écart au niveau scolaire et de notions communes), un indicateur *historique* qui prend en compte le fait que l'apprenant ait potentiellement déjà travaillé et plus ou moins bien réussi les chapitres candidats, et un indicateur *nouveauté* pour renouveler l'offre de recommandations en prenant en compte la fréquence et le rang de présentation des chapitres recommandés.

Pour l'étude des facteurs influençant le suivi des recommandations (QR1), les traces d'usage des apprenants collectées et analysées nous ont permis de constater que les recommandations sont davantage suivies lorsqu'elles interviennent à la suite d'un quiz échoué ou d'un chapitre terminé. Les recommandations associées à la stratégie de révision sont plus suivies que celles associées aux stratégies de continuité et d'approfondissement, avec un choix plus important des apprenants en faveur des chapitres de niveau scolaire équivalent au chapitre qui vient d'être étudié. Concernant les usages des recommandations (QR2), nous avons mis en évidence par l'analyse de traces, différents niveaux d'implication dans les chapitres recommandés suivis, allant de la simple consultation à l'étude approfondie des ressources du chapitre recommandé. Nous avons également constaté que les sessions de travail avec suivi de recommandations étaient plus productives en termes de nombre de ressources travaillées et de temps passé à étudier. La troisième question (QR3) s'intéressait aux leviers d'amélioration du système de recommandations identifiables par l'analyse des traces de navigation libre dans l'application. Outre les activités consécutives à un non-suivi des recommandations non prises en compte dans le principe de recommandation (étude d'une autre discipline, étude du même chapitre), ces analyses ont mis en évidence de probables lacunes dans le graphe de notions initial et la possibilité de redéfinir certaines stratégies de recommandation pour correspondre davantage aux choix des apprenants. Ces conclusions amènent donc plusieurs perspectives pour poursuivre l'étude.

9.1. AMÉLIORER LE MODÈLE DE RECOMMANDATION PAR UNE APPROCHE ITÉRATIVE

Les analyses d'usage du système de recommandations ont mis en évidence de potentiels leviers d'amélioration du modèle de recommandation. Nous avons pour objectif d'utiliser une approche itérative qui permettrait de modifier certains paramètres du modèle de recommandation et de proposer ces modèles alternatifs aux apprenants afin d'ajuster le modèle selon les usages observés.

Comme discuté dans la section 8, un premier modèle alternatif consisterait à laisser aux apprenants la possibilité d'aller explorer les recommandations des autres stratégies. Ce modèle alternatif plus ouvert que le modèle initial pourrait palier aux 5.2% et 25.5% des recommandations non-suivies au profit de chapitres non recommandés car non classés parmi les plus pertinents pour la stratégie assignée (voir Figure 12).

Les analyses ayant mis en évidence que les chapitres de niveaux scolaires équivalents au chapitre actuel étaient préférentiellement suivis au détriment des chapitres de niveaux scolaires distincts (voir Tableau 4), nous prévoyons également un modèle alternatif qui masquerait le niveau scolaire du chapitre recommandé pour confirmer l'hypothèse d'un biais

supposés sur l’interface d’affichage des recommandations.

Enfin, exploiter les liens découverts dans la navigation libre des apprenants pourrait permettre de renforcer le graphe de notions existant qui semble avoir ses limites comme discuté dans la section 8.

Le terrain expérimental industriel nous permet d’aller étudier ces hypothèses en confrontant ces différents modèles alternatifs à des populations d’apprenants distincts afin d’observer directement les effets de ces alternatives sur les usages.

9.2. AUGMENTER LA DIVERSITÉ DES RESSOURCES À RECOMMANDER

L’une des principales difficultés lors de la conception de notre système de recommandations a été le **faible nombre et le manque de diversité des ressources pédagogiques disponibles associées au graphe de notions**. En effet, le système ne pouvait recommander que des ressources internes à l’application Nomad Education et le graphe de notions a été conçu pour indexer des chapitres uniquement. Un apprenant désirant travailler en profondeur une thématique se trouvait donc rapidement confronté à des recommandations jugées trop éloignées de son chapitre actuel, les apprenants ne comprenant pas toujours l’intérêt d’étudier des chapitres en dehors de leur programme scolaire. Il serait intéressant d’appliquer ce même moteur de recommandations dans un contexte plus large avec un graphe de notions incluant un nombre de ressources pédagogiques plus important, pour travailler plus en profondeur une même thématique. La thèse de Connes (2023) portait précisément sur la recommandation de ressources éducatives à partir d’une large banque de ressources en s’appuyant sur des Ressources Éducatives Libres (REL) issues de différentes plateformes et indexées par un processus semi-automatique. Une telle banque de ressources permettrait d’avoir une diversité de ressources traitant de la même thématique et donc de répondre également aux retours des professeurs évaluateurs du système qui recommandaient de continuer à travailler sur le même chapitre mais avec des exercices différents.

Outre le nombre de ressources indexées dans le graphe de notions, il serait bénéfique de pouvoir proposer des ressources de natures différentes (exercices interactifs, *flashcards*, vidéos, annales d’examens). Cela permettrait d’affiner la granularité des recommandations proposées avec des suggestions de types de ressources précises et non plus de chapitres, ceci afin d’avoir des recommandations peut-être plus concrètes pour l’apprenant. Néanmoins, il conviendrait alors d’évaluer les apports pédagogiques de chaque type de ressources, toutes ne répondant pas à la même intention pédagogique (activité de découverte, d’apprentissage, d’auto-évaluation...).

Ceci impliquerait également d’ajouter un nouveau paramétrage dans le modèle de recommandation, relatif au type de contenu recommandé, similairement aux travaux menés par (Fazazi *et al.*, 2018) qui incluent les préférences des utilisateurs, leur stade d’apprentissage et l’apport pédagogique de la ressource pour choisir le type de ressources à inclure dans la personnalisation. Toutefois, avec ce nombre de ressources plus important ou plus diversifié, le système de recommandations actuel tendrait à proposer des ressources du même niveau scolaire, ce qui nous éloignerait de l’approche pédagogique défendue dans ces travaux, à savoir l’étude par notions transversales aux différents niveaux scolaires. Il faudrait donc paramétrer le système pour prendre en compte cette apparente contradiction et éventuellement choisir la politique la plus adaptée suivant l’apprenant.

10. REMERCIEMENTS

Le développement des modèles et les analyses des usages ont été réalisés dans le cadre d'une thèse CIFRE avec l'entreprise Nomad Education.

RÉFÉRENCES

- Allègre, O., Yessad, A., et Luengo, V. (2021). Amélioration de l'interprétabilité du diagnostic cognitif de l'apprenant par catégorisation des composantes de connaissance. *10e Conférence sur les Environnements Informatiques pour l'Apprentissage Humain*, 34-45.
- Averell, L., et Heathcote, A. (2011). The form of the forgetting curve and the fate of memories. *Journal of Mathematical Psychology*, 55(1), 25-35. <https://doi.org/10.1016/j.jmp.2010.08.009>
- Badier, A., Lefort, M., et Lefevre, M. (2023). Recommendation model for an after-school e-learning mobile application. *CSEDU*. <https://doi.org/10.5220/0011717800003470>
- Baker, R. (2001). *The basics of item response theory* (2nd ed). ERIC Clearinghouse on Assessment and Evaluation.
- Baker, R., Ma, W., Zhao, Y., Wang, S., et Ma, Z. (2020). The results of implementing zone of proximal development on learning outcomes. *The 13th International Conference on Educational Data Mining*.
- Bloom, B. (1956). *Taxonomy of educational objectives : the classification of educational goals*. Longmans.
- BO. (2020). Bulletin officiel n°31 du 30 juillet 2020. *Ministère de l'Éducation Nationale et de la Jeunesse*. https://www.education.gouv.fr/pid285/bulletin_officiel.html?pid_bo=39771.
- BO. (2022). Programmes et ressources en mathématiques - voie GT.
- Branthôme, M. (2022). Pyrates : a serious game designed to support the transition from block-based to text-based programming. *EC-TEL*, 13450, 31-44.
- Brusilovsky, P., et Peylo, C. (2003). Adaptive and intelligent web-based educational systems. *International Journal of Artificial Intelligence in Education*, 13, 156-169.
- Castells, P., Hurley, N., et Vargas, S. (2022). Novelty and diversity in recommender systems. Dans F. Ricci, L. Rokach et B. Shapira (dir.), *Recommender Systems Handbook* (p. 603-646). Springer US. https://doi.org/10.1007/978-1-0716-2197-4_16
- Choffin, B., Popineau, F., Bourda, Y., et Vie, J.-J. (2019). DAS3H : modeling student learning and forgetting for optimally scheduling distributed practice of skills. *Junior Conference on Data Science and Engineering*.
- Clément, B., Roy, D., Oudeyer, P.-Y., et Lopes, M. (2015). Multi-armed bandits for intelligent tutoring systems. *Journal of Educational Data Mining*, 7(2), 20-48.
- Connes, V. (2023). *Recommandation de ressources éducatives libres dans le projet X5GON*. [Thèse de doctorat, Nantes Université, Nantes, France]. <https://theses.hal.science/tel-04101102>
- Corbett, A. T., et Anderson, J. R. (1994). Knowledge tracing : modeling the acquisition of procedural knowledge. *Proceedings of UMUAI*, 253-278.
- Daher, J. B., Brun, A., et Boyer, A. (2018). Multi-Source data mining for e-learning. *7th International Symposium "From Data to Models and Back (DataMod)"*.
- De Maio, C., Fenza, G., Gaeta, M., Loia, V., Orciuoli, F., et Senatore, S. (2012). RSS-based e-learning recommendations exploiting fuzzy FCA for knowledge modeling. *Applied Soft Computing*, 12(1), 113-124.

- Décret. (2015). Décret n° 2015-372 du 31 mars 2015 relatif au socle commun de connaissances, de compétences et de culture.
- Ebbinghaus, H. (2013). Memory : a contribution to experimental psychology. *Annals of Neurosciences*, 20(4), 155-156.
- Education Nationale. (2021). Programmes d'enseignement de l'école élémentaire et du collège : cycles 2, 3 et 4 - data.gouv.fr. <https://www.data.gouv.fr/fr/datasets/programmes-denseignement-de-lecole-elementaire-et-du-college-cycles-2-3-et-4/>
- El-Kechaï, N., Melero, J., et Labat, J.-M. (2015). Adaptation de serious games selon la stratégie choisie par l'enseignant : approche fondée sur la compétence-based knowledge space theory. *7ème Conférence sur les Environnements Informatiques pour l'Apprentissage Humain (EIAH 2015)*, 294.
- Elodie, P., et Monseur, C. (2018). La multidimensionnalité dans l'évaluation PISA 2003 : la place d'une dimension générale. *e-JIREF*, 4(3), 21-45.
- Erdt, M., Fernandez, A., et Rensing, C. (2015). Evaluating recommender systems for technology enhanced learning : a quantitative survey. *IEEE Transactions on Learning Technologies*, 326-344.
- Fazazi, H. E., Qbadou, M., Salhi, I., et Mansouri, K. (2018). Personalized recommender system for e-learning environment based on student's preferences. *IJCSNS International Journal of Computer Science and Network Security*, 18(10), 6.
- Guruge, D. B., Kadel, R., et Halder, S. J. (2021). The state of the art in methodologies of course recommender systems—a review of recent research. *Data*, 6(2), 18. <https://doi.org/10.3390/data6020018>
- Hubert, N., Monnin, P., Brun, A., et Monticolo, D. (2022). New strategies for learning knowledge graph embeddings : the recommendation case. https://doi.org/10.1007/978-3-031-17105-5_5
- Jiang, W., Pardos, Z. A., et Wei, Q. (2019). Goal-based course recommendation. *Proceedings of the 9th International Conference on Learning Analytics and Knowledge*, 36-45.
- Juhel, J. (1999). Étude de la dimensionnalité d'un test de raisonnement l'aide des modèles de réponse à l'item. *Psychologie & Psychométrie*, 20, 85-111.
- Kizilcec, R. F., Piech, C., et Schneider, E. (2013). Deconstructing disengagement : analyzing learner subpopulations in massive open online courses. *Proceedings of the Third International Conference on Learning Analytics and Knowledge*, 170-179. <https://doi.org/10.1145/2460296.2460330>
- Kotkov, D., Wang, S., et Veijalainen, J. (2016). A survey of serendipity in recommender systems. *Knowledge-Based Systems*, 111, 180-192. <https://doi.org/10.1016/j.knosys.2016.08.014>
- Krauss, C., Chandru, R., Merceron, A., An, T.-S., Zwicklbauer, M., et Arbanowski, S. (2016). You might have forgotten this learning content ! How the smart learning recommender predicts appropriate learning objects. *International Journal on Advances in Intelligent Systems*, 9 (3-4), 472-484.
- Lü, L., Medo, M., Yeung, C. H., Zhang, Y.-C., Zhang, Z.-K., et Zhou, T. (2012). Recommender systems. *Physics Reports*, 519, 1-49. <https://doi.org/10.1016/j.physrep.2012.02.006>
- Murray, T. (2003). Eon : authoring tools for content, instructional strategy, student model and interface design. Dans T. Murray, S. B. Blessing et S. Ainsworth (dir.), *Authoring Tools for Advanced Technology Learning Environments : Toward Cost-Effective Adaptive, Interactive and Intelligent Educational Software* (p. 309-339). Springer Netherlands.

- Nguyen, C., Roussanaly, A., et Boyer, A. (2014). Learning resource recommendation : an orchestration of content-based filtering, word semantic similarity and page ranking. *9th European Conference on Technology Enhanced Learning, EC-TEL 2014*.
- Nikolayeva, I., Martin, B., Yessad, A., Chenevotot, F., Pilet, J., Prévité, D., Grugeon-Allys, B., et Luengo, V. (2018). How to help teachers adapt to learners? Teachers' perspective on a competency and error-Type centered dashboard. *13th EC-TEL*, 596-599.
- Nikou, S., et Economides, A. (2018). Mobile-based micro-learning and assessment : impact on learning performance and motivation of high school students. *Journal of Computer Assisted Learning*, 34(3), 269-278. <https://doi.org/10.1111/jcal.12240>
- Ortegón Romero, O. Y., et Krug Wives, L. (2021). Systematic review of context-aware systems that use item response theory in learning environments. *Adjunct Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization*, 100-104. <https://doi.org/10.1145/3450614.3464482>
- Paquette, G., Marino, O., Rogozan, D., et Léonard, M. (2015). Competency-based personalization for massive online learning. *Smart Learning Environments*, 2. <https://doi.org/10.1186/s40561-015-0013-z>
- Pinsky, L. E., et Irby, D. M. (1997). "If at first you don't succeed" : using failure to improve teaching. *Academic Medicine : Journal of the Association of American Medical Colleges*, 72(11), 973-976, discussion 972. <https://doi.org/10.1097/00001888-199711000-00013>
- Plaskura, P. (2019). Modelling of forgetting curves in educational e-environment. *Information Technologies and Learning Tools*, 71(3), 1-11. <https://doi.org/10.33407/itlt.v71i3.2368>
- Reich, J. (2014). MOOC completion and retention in the context of student intent. *Educause Review*.
- Rizun, M. (2019). Knowledge graph application in education : a literature review. *Acta Universitatis Lodzianis. Folia Oeconomica*, 3(342), 7-19. <https://doi.org/10.18778/0208-6018.342.01>
- Sablaylorles, L., Lefevre, M., Guin, N., et Broisin, J. (2022). Design and evaluation of a competency-based recommendation process. *Proceedings of ITS*.
- Urda-neta-Ponte, M. C., Mendez-Zorrilla, A., et Oleagordia-Ruiz, I. (2021). Recommendation systems for education : systematic review. *Electronics*, 10(14). <https://doi.org/10.3390/electronics10141611>
- Venant, R., Sharma, K., Dillenbourg, P., Vidal, P., et Broisin, J. (2017). A study of learners' behaviors in hands-on learning situations and their correlation with academic performance. *AIED*, 10331.
- Vie, J.-J. (2016). Modèles de tests adaptatifs pour le diagnostic de connaissances dans un cadre d'apprentissage à grande échelle. Récupérée juin 29, 2023, à partir de <https://www.theses.fr/2016SACL090>
- Vie, J.-J., et Kashima, H. (2019). Knowledge tracing machines : factorization machines for knowledge tracing. *Proceedings of the AAAI Conference on AI*.
- Vlach, H. A., et Sandhofer, C. M. (2012). Distributing learning over time : the spacing effect in children's acquisition and generalization of science concepts. *Child Development*, 83(4), 1137-1144.
- Vygotskiï, L. S., et Cole, M. (1978). *Mind in society : the development of higher psychological processes*. Harvard University Press.
- Wauters, K., Desmet, P., et den Noortgate, W. V. (2021). Adaptive item-based learning environments based on the item response theory : possibilities and challenges. *Journal of Computer Assisted Learning*, 26(6), 549-562.

- Yanes, N., Mohamed Mostafa, A., Ezz, M., et Naif Almuayqil, S. (2020). A machine learning-based recommender system for improving students learning experiences. *IEEE Access*, 8, 201218-201235.
- Zhao, Z., Hong, L., Wei, L., Chen, J., Nath, A., Andrews, S., Kumthekar, A., Sathiamoorthy, M., Yi, X., et Chi, E. (2019). Recommending what video to watch next : a multitask ranking system. *Proceedings of the 13th ACM Conference on Recommender Systems*, 43-51.
- Ziarani, R. J., et Ravanmehr, R. (2021). Serendipity in recommender systems : a systematic literature review. *Journal of Computer Science and Technology*.

Comités

REDACTRICE EN CHEF

Élise LAVOUÉ • LIRIS, Université Jean Moulin Lyon 3

COMITES DE REDACTION

Monique BARON • LIP6, Sorbonne Université, Paris

Laetitia BOULC'H • EDA, Université Paris Cité

Julien BROISIN • IRIT, Université de Toulouse Paul-Sabatier

Pierre-André CARON • CIREL, Université de Lille

Michel DESMARAIS • Polytechnique Montréal, Canada

Philippe DESSUS • LSE, Université Grenoble Alpes

Béatrice DROT-DELANGE • ACTé, Université Clermont Auvergne

Nour EL MAWAS • CIREL, Université de Lille

Sébastien GEORGE • LIUM, Le Mans Université, Laval

Élise LAVOUÉ • LIRIS, Université Jean Moulin Lyon 3

Vanda LUENGO • LIP6, Sorbonne Université, Paris

Agathe MERCERON • Université de Berlin, Allemagne

Gaëlle MOLINARI • TECFA, Unidistance, Genève, Suisse

Chrysta PÉLISSIER • Praxiling, Université Montpellier 3

Gaëtan TEMPERMAN • Université de Mons, Belgique

COMITE DE LECTURE

François-Xavier BERNARD • EDA, Université Paris Cité

Mireille BÉTRANCOURT • TECFA, Université de Genève, Suisse

Jacques BÉZIAT • CIRNEF, Université de Caen Normandie

Bernard BLANDIN • CREF, Université Paris Nanterre et CESI

François BOUCHET • LIP6, Sorbonne Université

Éric BRUILLARD • EDA, Université Paris Cité

Thibault CARRON • LIP6, Sorbonne Université et Université de Savoie Mont-Blanc

Ullrich CARSTEN • EdTec Lab, DFKI GmbH, Sarrebrück, Allemagne

Ghislaine CHARTRON • CNAM, Paris

Jacques CRINON • INSPE, Université Paris Est Créteil

Nicolas DELESTRE • LITIS, INSA de Rouen

Cyrille DESMOULINS • LIG, Université Grenoble Alpes

Angélique DIMITRACOPOULOU • LTEE, Université d'Egée, Grèce
Hassina EL KECHAI • TECHNE, Université de Poitiers
Stéphanie FLECK, PErSEUs • Université de Lorraine – INSPE
Cédric FLUCKIGER • Théodile-CIREL, Université Lille
Michel GALAUP • EFTS, Université Toulouse 2
Serge GARLATTI • Lab-STICC, IMT Atlantique, Brest
Jean-Marie GILLIOT • Lab-STICC, IMT Atlantique, Brest
Monique GRANDBASTIEN • LORIA, Université de Lorraine
Brigitte GRUGEON • LDAR, INSPE, Université Paris Est Créteil
Nicolas GUICHON • ICAR, Université Lumière Lyon 2
Nathalie GUIN • LIRIS, Université Lyon 1
France HENRI • LICEF, Télé-université, Université du Québec, Montréal, Canada
Richard HOTTE • LICEF, Télé-université, Université du Québec, Montréal, Canada
Céline JOIRON • MIS, Université de Picardie Jules Verne, Amiens
Mehdi KHANEBOUBI • STEF, ENS Paris-Saclay
Vassilis KOMIS • Université de Patras, Grèce
Thérèse LAFERRIÈRE • TACT, Université Laval, Canada
Pierre LAFORCADE, LIUM • Université du Mans
Marie LEFÈVRE • LIRIS, Université Lyon 1
Dominique LENNE • Heudiasyc, Université de Technologie de Compiègne
Pascal LEROUX • CREN, Le Mans Université
Paul LIBBRECHT • Leibniz Institute for Research and Information in Education,
Allemagne
Cabral LIMA • Université Fédéral de Rio de Janeiro, Brésil
Domitile LOURDEAUX • Heudiasyc, Université de Technologie de Compiègne
Nadine MANDRAN, LIG • Université de Grenoble Alpes
Pascal MARQUET • LISEC, Université de Strasbourg
André MAYERS • Université de Sherbrooke, Canada
Christine MICHEL • TECHNÉ, Université de Poitiers
Nadia NAFFI • Université Laval
Roger NKAMBOU • GDAC, Université du Québec à Montréal, Canada
Thierry NODENOT • LIUPPA, Université de Pau et des Pays de l'Adour, Bayonne
Sandra NOGRY • Paragraphe, Cergy-Paris Université - INSPé de l'Académie de Versailles
Luc PAQUETTE • Université de l'Illinois à Urbana-Champaign, USA
Yvan PETER • LIFL, Université Lille, Villeneuve d'Ascq
Julia PILET • LDAR, ESPÉ, Université Paris Est Créteil Val de Marne
Valéry PSYCHÉ • Université TÉLUQ, Canada
Dominique PY • LIUM, Le Mans Université
Christophe REFFAY • ELLIAD, ESPÉ, Université de Franche-Comté
Éric SANCHEZ • TECFA, Université de Genève, Suisse
Karim SEHABA • LIRIS, Université Lyon 2

Nicolas SZILAS • TECFA, Université de Genève, Suisse
André TRICOT • EPSYLON, Université Paul-Valéry Montpellier 3
Rémi VENANT • LIUM, Le Mans Université, Laval
Kalina YACEF • Université de Sydney, Australie
Amel YESSAD • MOCAH-LIP6, Sorbonne Université

EN MEMOIRE D'ANCIENS MEMBRES DU COMITE

Erik DUVAL • Université de Louvain, Belgique
Jacques PERRIAULT • Université Paris Nanterre
François VILLEMONTAIX • CIREL, Université de Lille

NOUS REMERCIONS LES PERSONNES EXTERIEURES AUX COMITES QUI ONT RELU POUR CE VOLUME :

Marie-Hélène ABEL • HEUDIASYC, Université de Technologie de Compiègne
Laëtitia PIERROT • CREN, Le Mans Université
Bertrand MARNE • LIUM, Le Mans Université



ISBN 978-2-901384-09-0

DOI: 10.23709/sticef.31.1