

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 BADIÉ¹; 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.

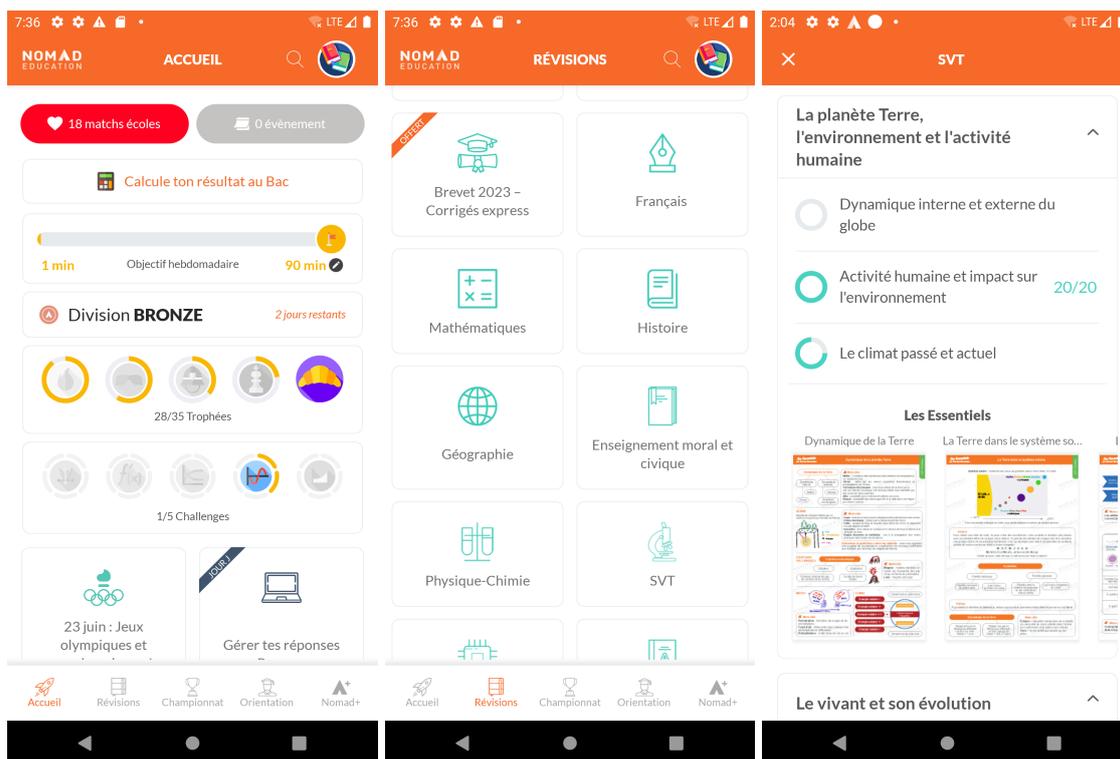


Figure 1 : Captures d'écran de l'application Nomad Education : à gauche la page "Accueil" commune à tous les niveaux, au milieu la page "Révisions" pour un niveau 3^{ème}, à droite un aperçu de la discipline SVT pour un niveau 3^{ème}

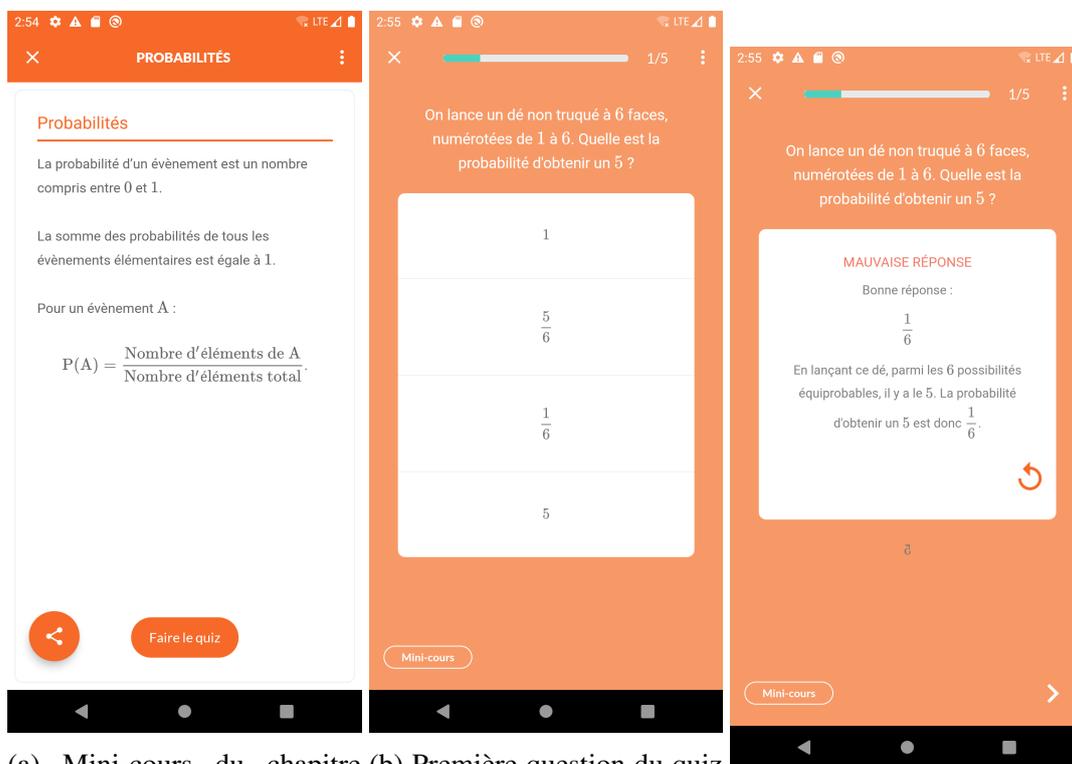
2.2. RESSOURCES DISPONIBLES DANS L'APPLICATION NOMAD EDUCATION

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'attendu 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.

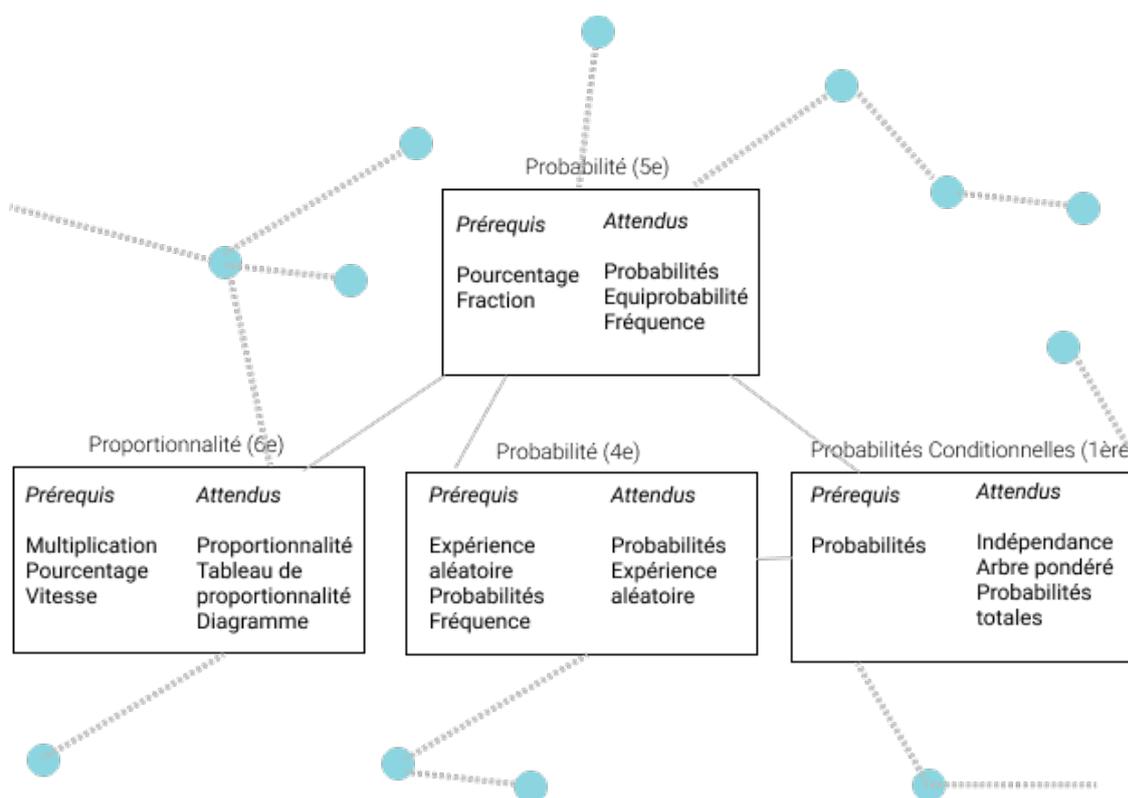


Figure 3 : Extrait du graphe de notions de l'application Nomad Education pour la discipline Mathématiques montrant un focus sur le chapitre *Probabilité 5^{ème}* et ses liens avec les autres chapitres de l'application, et illustrant en arrière plan, que ces autres chapitres ont également d'autres connexions

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).

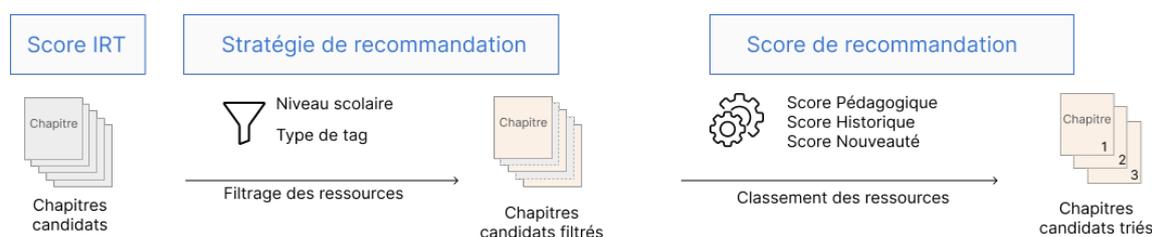


Figure 4 : Principe du modèle de recommandations pour une application mobile parascolaire

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.

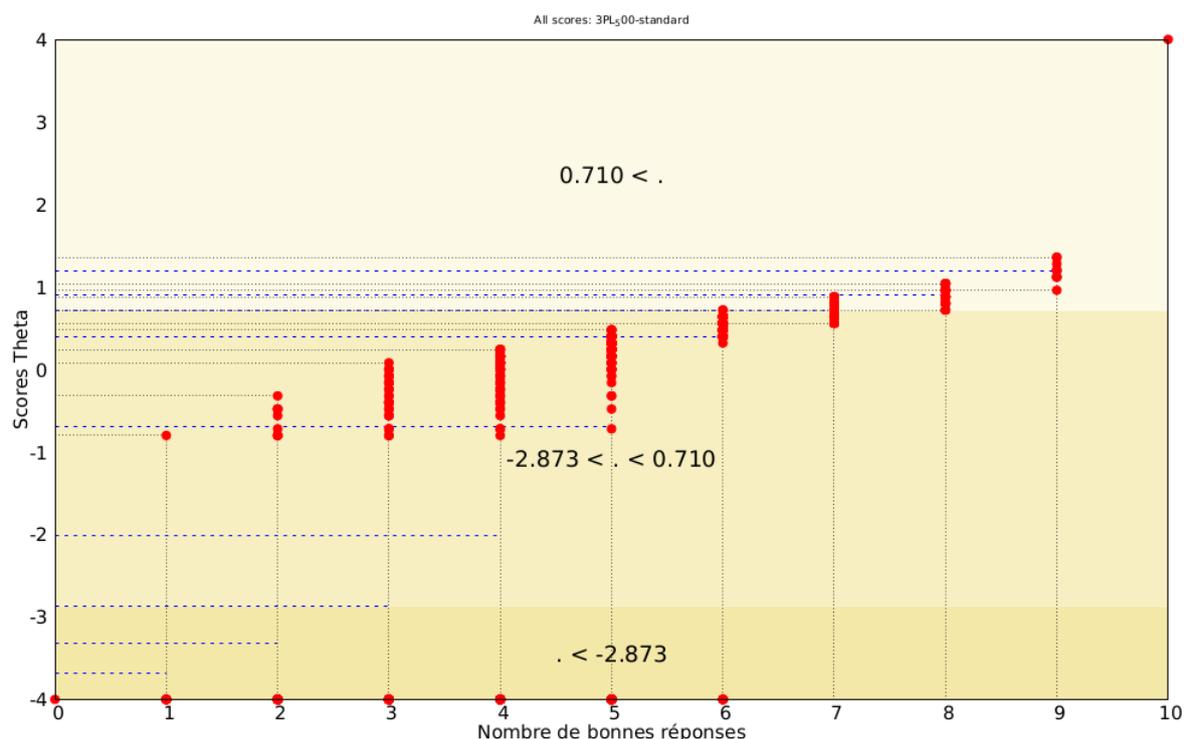


Figure 5 : Répartition des scores θ possibles selon le nombre de réponses correctes pour le quiz *Puissances 3^{ème}* et mise en évidence des seuils IRT. Chaque point rouge correspond à une valeur de θ pour une séquence S. Les pointillés bleus correspondent à la moyenne des θ observés pour les séquences 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.

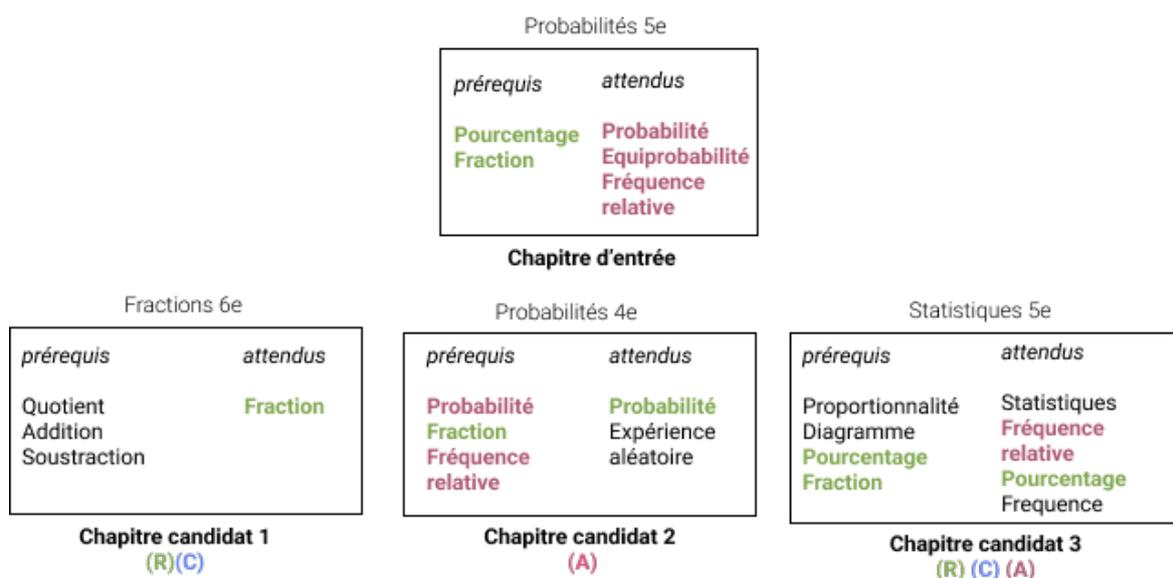


Figure 6 : Exemple de filtrage du graphe de notions selon la stratégie de recommandation sélectionnée. Les notions en commun avec les prérequis du chapitre en cours d'étude sont représentées en vert, celles en commun avec les attendus en rouge. Les chapitres marqués (R),(C) et (A) sont respectivement 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 débloquer 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.

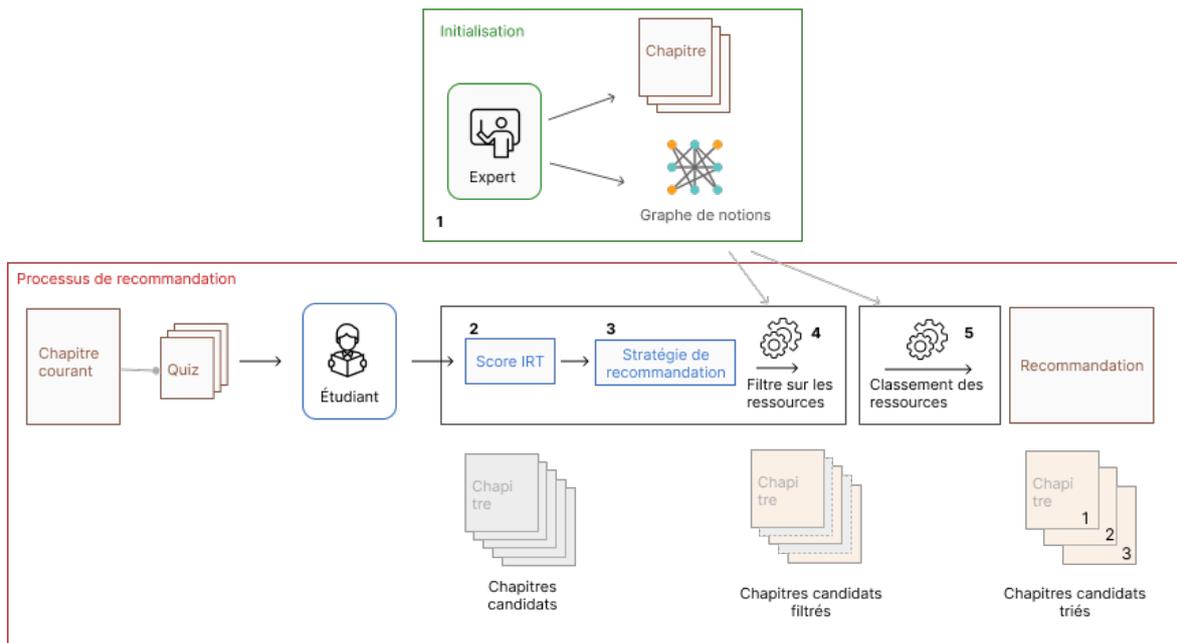


Figure 7 : Processus d'implémentation du système de recommandations

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).

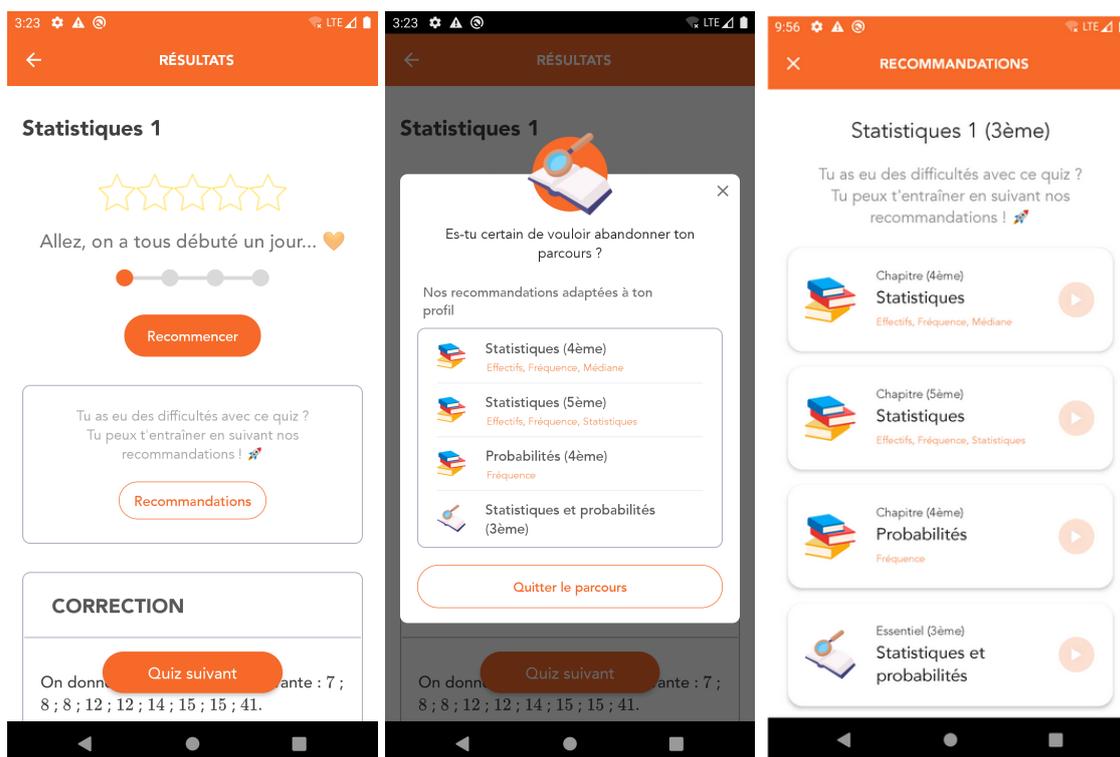


Figure 8 : Interfaces de recommandations pour le chapitre *Statistiques* de niveau 3^{ème} pour la stratégie révision. À gauche, par le point d'entrée quiz échoué, au milieu le cas d'un parcours interrompu et à droite l'interface de fin du parcours. Le niveau scolaire des chapitres recommandés est affiché 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.

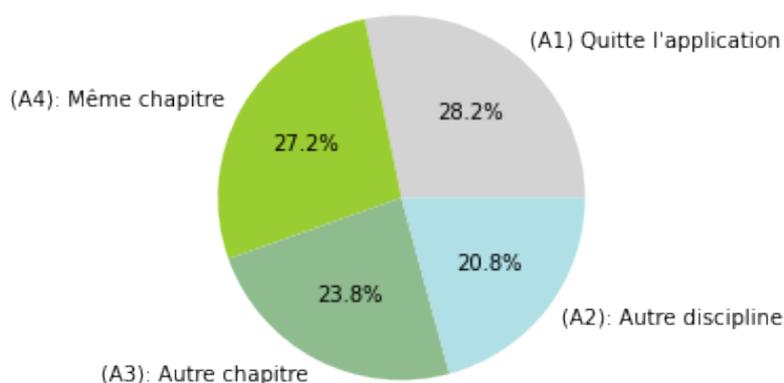


Figure 9 : Activités réalisées après un non-suivi des recommandations, d'après l'analyse des *AppEvents* (N=13 864)

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.

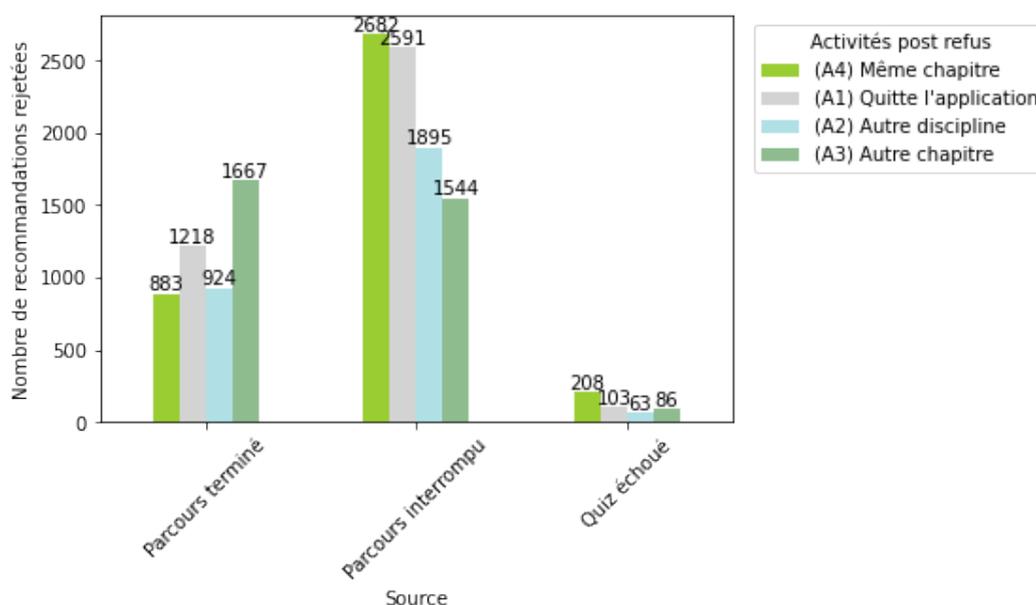


Figure 10 : Répartition des activités réalisées après un non-suivi des recommandations, selon la source de la recommandation vue

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

souhaits 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).

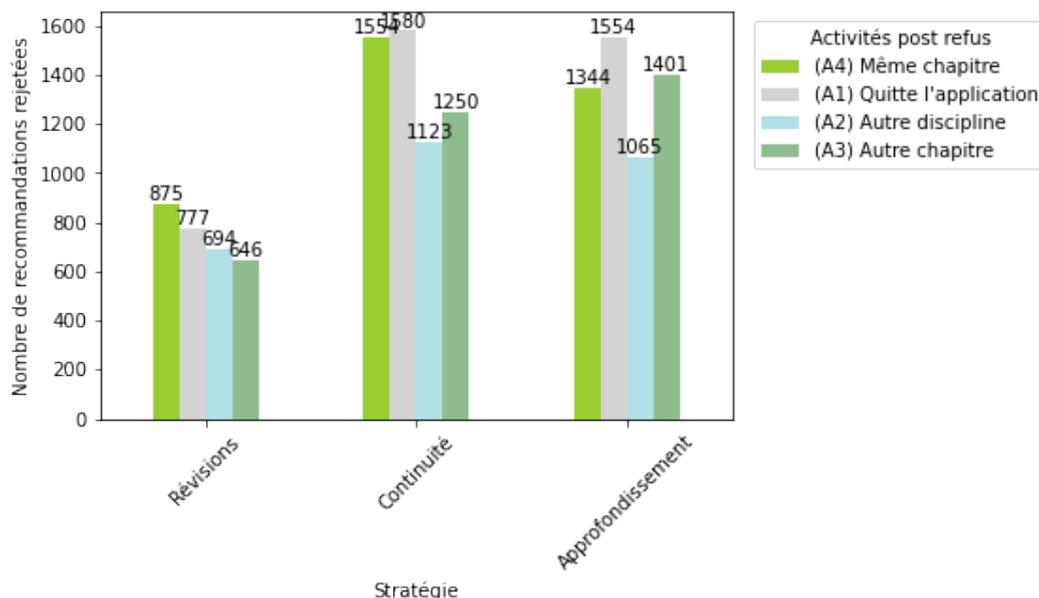


Figure 11 : Répartition des activités réalisées après non-suivi des recommandations selon la stratégie de la recommandation vue

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.

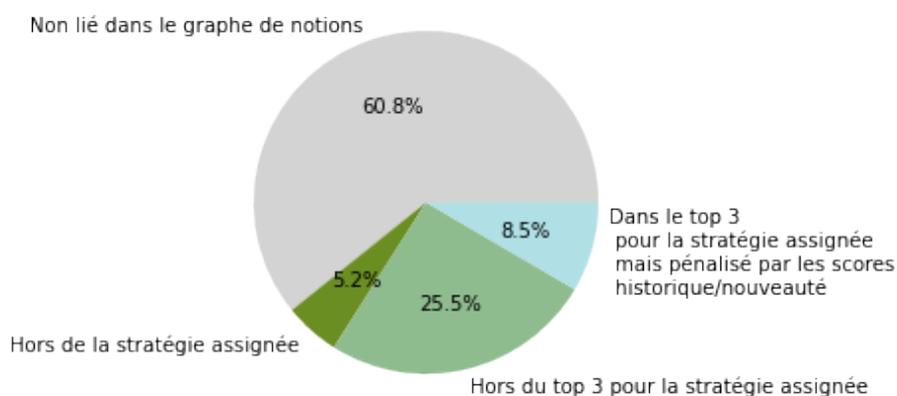


Figure 12 : Raisons pour lesquelles le système n'a pas recommandé le chapitre choisi par l'apprenant dans le cas des recommandations non-suivies avec activité A3

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 Lodzianae. 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*.