

# Sticef

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

**Volume 27, numéro 2, 2020**

*numéro spécial*

**Sélection  
de la conférence EIAH 2019**

*sous la direction de  
Julien BROISIN,  
Vanda LUENGO  
et Éric SANCHEZ*





*Sticef*



# *Sticef*

**Volume 27**  
**numéro 2, 2020**

*numéro spécial*  
**Sélection de la**  
**conférence**  
**EIAH 2019**

© ATIEF, 2020

ISBN 978-2-901384-04-5

DOI: 10.23709/sticef.27.2 en ligne sur [www.sticef.org](http://www.sticef.org)

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



## Sommaire

---

---

**Julien BROISIN, Vanda LUENGO, Éric SANCHEZ • Éditorial  
du numéro spécial : « Sélection de la conférence EIAH 2019 » 7**

### *Articles de Recherche*

**Yvan PETER, Yann SECQ, Marielle LÉONARD • Reconnaissance  
de motifs redondants et répétitions :  
introduction à la pensée informatique ..... 11**

**Fatima HARRAK, François BOUCHET, Vanda LUENGO • Liens  
entre performance, assiduité et questions posées  
et/ou questions votées en ligne dans le cadre  
d'une classe inversée..... 35**

**Roger NKAMBOU, Ange TATO, Janie BRISSON, Serge ROBERT,  
Maxime SAINTE-MARIE • Une approche hybride  
à la modélisation de l'apprenant dans un STI  
pour l'apprentissage du raisonnement logique..... 63**

**Olivier CATTEAU, Anne MAYERE, Nicolas SAVY,  
Odile BEYNE-RAUZY • Équiper et évaluer un  
« certificat de compétences cliniques »..... 103**

**Bertrand MARNE • Fonder un jeu sérieux sur un  
jouet-micromonde ..... 131**

**Comités ..... 163**







# Éditorial du numéro spécial : « Sélection de la conférence EIAH 2019 »

► **Julien BROISIN** (IRIT, Université Toulouse 3 Paul Sabatier),  
**Vanda LUENGO** (LIP6 MOCAH, Sciences Sorbonne Université),  
**Éric SANCHEZ** (TECFA, Université de Genève)

---

---

## 1. Introduction

Ce numéro spécial de la revue *Sticef* présente une sélection de cinq articles qui constituent des versions développées et révisées de papiers présentés lors de la 9<sup>e</sup> édition de la conférence EIAH (Environnements Informatiques pour l'Apprentissage Humain) qui s'est déroulée à Paris (France) du 4 au 7 juin 2019 et dont la thématique était : « Données numériques et prise en compte de l'apprenant dans les Environnements Informatiques pour l'Apprentissage Humain (EIAH) ».

Organisée tous les deux ans, cette conférence francophone, parrainée par l'Association des Technologies de l'Information pour l'Éducation et la Formation (ATIEF), a pour objectif la diffusion de travaux de recherche et la formation des chercheurs en lien avec les EIAH. Cette communauté pluridisciplinaire regroupe des chercheurs en informatique, en sciences et techniques de l'information et de la communication et en sciences humaines et sociales (sciences de l'éducation, didactique, psychologie, sciences de l'information et de la communication, etc.). Elle conduit des recherches dont la finalité est de problématiser et modéliser des environnements informatiques, mais aussi d'analyser les usages du numérique pour l'éducation et la formation afin d'en comprendre les effets. Ainsi, les travaux qui sont menés, souvent en collaboration étroite avec les praticiens du domaine, visent à produire des résultats de recherche, des modèles et des environnements informatiques dédiés à la conception, la mise en œuvre et l'évaluation de situations d'apprentissage instrumentées avec le numérique. Ces situations d'apprentissage peuvent se dérouler dans différents contextes, pour des publics divers (scolaire,

universitaire, formation professionnelle, public à besoins particuliers) et selon des modalités variées (en présence, à distance ou en mode hybride). Les thématiques traitées concernent la conception, les usages, l'adaptation, les interactions, les fondements technologiques, les contextes d'apprentissage et les questions éthiques et de protection des données.

## **2. Présentation du numéro spécial**

La thématique scientifique de la conférence EIAH 2019 était la question des données numériques et la manière dont elles peuvent être exploitées pour favoriser la prise en compte de l'apprenant dans les environnements numériques dédiés à l'éducation. Les cinq articles sélectionnés dans ce numéro spécial attestent de la diversité, de la vitalité et de la richesse de la recherche sur les EIAH.

L'article rédigé par Yvan Peter, Marielle Léonard et Yann Secq, intitulé « Reconnaissance de motifs redondants et répétitions : introduction à la pensée informatique », présente une expérimentation conduite auprès d'élèves de l'enseignement primaire dans le cadre d'un enseignement de la pensée informatique. L'objectif était de développer la capacité de reconnaissance de motifs et leur expression sous la forme de répétitions via des activités déconnectées et en ligne. Les analyses conduites sur les traces d'activité permettent d'attester de l'acquisition des compétences visées.

Dans leur article « Liens entre performance, assiduité et questions posées et/ou questions votées en ligne dans le cadre d'une classe inversée », Fatima Harrak, François Bouchet et Vanda Luengo s'intéressent aux questions posées par des étudiants de première année de médecine, aux votes exprimés sur ces questions et à leur utilisation par les professeurs. L'annotation automatique appliquée à l'ensemble du corpus et la comparaison entre étudiants votants et non-votants révèlent que le fait de poser des questions est associé à un apprentissage plus important et montre la valeur du vote comme alternative à la formulation de questions pour les étudiants sachant déjà formuler leurs propres questions.

Ce sont les questions de la modélisation de l'apprenant et de la conception d'un tuteur intelligent qui sont abordées par Roger Nkambou, Ange Tato, Janie Brisson, Serge Robert et Maxime Sainte-Marie dans leur article intitulé « Une approche hybride à la modélisation de l'apprenant dans un STI pour l'apprentissage du raisonnement logique ». Ce système a été conçu pour favoriser le développement des compétences en raisonnement logique. Les travaux qui ont été menés ont permis de

développer et de valider un réseau bayésien qui permet l'inférence du comportement de l'apprenant. L'approche qui est proposée constitue une approche hybride de la modélisation de l'apprenant.

L'article « Équiper et évaluer un “certificat de compétences cliniques?” » rédigé par Olivier Catteau, Anne Mayere, Nicolas Savy et Odile Beyne-Rauzy, s'intéresse à la question de la certification des connaissances. Les auteurs ont conçu, développé et expérimenté un environnement informatique qui permet d'instrumenter un examen pour l'obtention d'un certificat de compétences cliniques. L'environnement proposé permet, en particulier, de recueillir les données des évaluations, d'automatiser et d'accélérer leurs traitements. Les réponses aux questionnaires diffusés auprès des étudiants et des enseignants montrent l'intérêt de la démarche. Les analyses statistiques permettent de conclure à l'absence de redondance entre cette forme d'examen et les examens théoriques conduits par ailleurs.

Enfin, dans son article intitulé « Fonder un jeu sérieux sur un jouet-micromonde », Bertrand Marne aborde la question de la conception des jeux sérieux. À travers trois études de terrain, ses travaux visent à évaluer un ensemble de méthodes de conception afin d'identifier des patrons de conception de jeux sérieux. Elles lui ont permis d'identifier plusieurs verrous liés à la conception de jeux sérieux articulés sur les jouets systémiques et de proposer des solutions pour les lever. En particulier, un patron de conception pour la création de jeux sérieux est suggéré, qui reste à affiner et à tester.





# Reconnaissance de motifs redondants et répétitions : introduction à la pensée informatique

► **Yvan PETER** (Univ. Lille, CNRS, Centrale Lille, UMR 9189 CRISAL, France), **Yann SECQ** (Univ. Lille, CNRS, Centrale Lille, UMR 9189 CRISAL, France), **Marielle LÉONARD** (Univ. Lille, CNRS, Centrale Lille, UMR 9189 CRISAL, France ; CIREL, Centre Interuniversitaire de Recherche en Éducation de Lille, EA 4354)

---

---

■ **RÉSUMÉ** • Cet article présente une expérimentation d'initiation d'élèves de primaire à la pensée informatique. L'objectif est de développer la capacité de reconnaître des motifs et de les exprimer sous la forme de répétitions via des activités déconnectées et en ligne. Les résultats montrent un impact positif des activités ainsi que les compétences acquises.

■ **MOTS-CLÉS** • Pensée informatique, programmation, séquence pédagogique, école primaire.

■ **ABSTRACT** • *This article presents an experimentation to introduce Computational Thinking to primary school pupils. The aim is to enhance their capability to spot repetitive patterns and to express them as loops through unplugged and plugged-in activities. The results show a positive impact of the activities and the competencies acquired.*

■ **KEYWORDS** • *Computational thinking, programming, pedagogical sequence, primary school.*

## **1. Introduction**

La pensée informatique (PI) correspond à des compétences d'abstraction et de résolution de problèmes qui peuvent être exploitées dans de nombreux contextes et appliquées à des domaines divers. La PI en tant que compétence cognitive ne se réduit pas à la programmation. Toutefois, elle vise à exprimer les choses d'une manière qui se prête à un traitement informatique. En ce sens, elle permet d'aborder la dimension algorithmique et la programmation en général. Dans son article fondateur, J. Wing (2006) appelle à l'enseignement de la pensée informatique pour tous et assez tôt dans la scolarité. Depuis, celle-ci a été appliquée à de nombreux domaines tels que les mathématiques et les sciences expérimentales (Weintrop *et al.*, 2016), les arts (Knochel et Patton, 2015) et même l'apprentissage des langues (Sabitzer *et al.*, 2018).

Cet article présente notre travail sur l'acquisition de compétences de base de la PI avec des élèves d'école primaire (8-10 ans). Nous nous sommes focalisés sur la capacité des élèves à reconnaître des motifs répétitifs et à les exprimer sous forme de répétitions. Cette abstraction et cette expression ouvrent la voie à un traitement systématique des données.

Nous présentons dans cet article les résultats de l'expérimentation menée en 2018. Celle-ci a impliqué 20 classes venant de 16 écoles pour un total de 447 élèves. Nous avons fait passer aux élèves des questionnaires avant et après l'activité pour mesurer leur capacité à identifier et à coder les répétitions. Nous avons également collecté les traces d'activité de programmation sur l'exerciseur en ligne.

Dans la suite, nous passons en revue l'état de l'art concernant la PI, comment elle est introduite à l'école et comment est réalisée l'évaluation des compétences associées. La section 3 présente la séquence pédagogique que nous avons définie et le contexte expérimental. Nous discuterons l'analyse des résultats dans la section 4 avant de conclure.

## **2. État de l'art**

### **2.1. La pensée informatique**

Il n'existe pas de définition unanimement acceptée du terme pensée informatique. Celle-ci se réfère à un certain nombre d'habilités (abstraction, réflexion algorithmique...) qui ouvrent la voie à un traitement automatisé. On pourra se référer à ce sujet au compte-rendu de la table ronde qui a eu lieu au colloque Didapro 7 (Drot-Delange *et al.*, 2019).

Les premiers concepts de PI datent des travaux de Seymour Papert avec le langage Logo (1972). Plus récemment, l'article de Jeannette Wing présentant la PI comme une compétence de base au même titre que la lecture, l'écriture et l'arithmétique a déclenché un vif intérêt au sein des communautés éducatives et de recherche (Wing, 2006). Wing insiste notamment sur le fait que la PI ne se réduit pas à la programmation mais correspond plutôt à la capacité de manipuler des abstractions et de résoudre les problèmes qui peuvent être appliqués à différents domaines. Elle prône l'apprentissage de ces compétences dès l'école.

Depuis, un nombre croissant de travaux de recherche se sont développés pour explorer les moyens d'introduire la PI à l'école. Quels sont les concepts fondamentaux à enseigner ? Quelles technologies se prêtent le mieux à ces apprentissages ? etc. Ces questions sont d'autant plus importantes que de nombreux pays ont commencé à introduire la PI et plus généralement « le code » dans les programmes à différents niveaux scolaires.

Rich *et al.* (2017) ont défini des *Trajectoires d'Apprentissage* en se basant sur une étude de la littérature. Une *Trajectoire d'Apprentissage* définit les concepts qui peuvent être abordés en fonction de la classe et avec quel niveau de détails. Elle est formalisée par un ensemble d'objectifs d'apprentissage associés à un chemin d'apprentissage permettant la réalisation de ces objectifs ainsi qu'à des exemples d'activités qui peuvent être mises en œuvre. Les auteurs relèvent que la plupart des travaux sont relatifs à un seul objectif d'apprentissage ou que, quand il y en a plusieurs, ils sont indépendants. Ils observent également que ces objectifs ont pu être appliqués à différents niveaux dans la mesure où ils s'adressent en général à des novices. Pour cette raison, ils se sont inspirés des approches pédagogiques et des programmes de mathématiques pour définir l'ordre d'introduction des concepts de la PI (chemin d'apprentissage). L'article illustre leur approche sur les concepts de Séquence, Répétition et Conditionnelle.

Ching *et al.* (2018) ont réalisé une étude des technologies disponibles pour l'enseignement de la PI. Ils ont identifié *les robots pédagogiques, les kits robotiques, les jeux de plateau, les applications en réalité augmentée, les applications ou sites Web pour la programmation* (graphique ou textuelle), *les outils de développement orientés animation ou jeu*. Ces différentes catégories se distinguent selon qu'elles utilisent une manipulation physique ou une interaction sur l'écran et selon que la rétroaction est concrète (par ex.,

mouvement d'un robot) ou abstraite (c.-à-d. retour sur l'écran). Les résultats de cette étude montrent que les concepts abordés avec ces technologies vont des séquences et répétitions à des concepts plus avancés et peuvent aussi impliquer des activités créatives ou de résolution de problèmes.

Gouws *et al.* (2013) proposent un cadre de description des compétences liées à la PI. En se basant sur une revue de la littérature, ils définissent un ensemble de compétences qui peuvent être abordées à travers la programmation : *Processus et Transformations, Modèles et Abstractions, Motifs et Algorithmes, Outils et Ressources, Inférence et Logique, Évaluations et Améliorations*. Ils combinent ces compétences avec des niveaux de maîtrise inspirés de la taxonomie de Bloom : *Reconnaître, Comprendre, Appliquer et Assimiler*. Ce cadre vise à être utilisé pour l'analyse et pour la conception d'activités.

Weintrop *et al.* (2016) s'intéressent à l'introduction des pratiques de la PI en mathématiques et en sciences de manière à permettre une définition des activités de PI indépendante de l'informatique. Les auteurs définissent une taxonomie de 22 pratiques groupées selon les catégories suivantes : *Données et Informations, Modélisation et Simulation, Calcul, Résolution de problèmes et Approche Systémique*.

Nos travaux s'inscrivent dans les compétences associées aux motifs et algorithmes élaborées par Gouws *et al.* qui associent reconnaissance des similarités ou motifs et leur traitement algorithmique (la répétition étant un des concepts mobilisables). Nous avons également une approche similaire à Rich *et al.* qui cherchent dans leurs trajectoires d'apprentissage à introduire les concepts par des exemples hors informatique et/ou par des activités débranchées. Ces auteurs ont notamment travaillé sur la répétition. Un des objectifs d'apprentissage relevés est de reconnaître la nécessité d'une structure répétitive avant même d'aborder la notion de répétition en programmation. Notre approche rentre dans la catégorie des jeux de plateau (déplacement d'un personnage sur une grille en débranché et en numérique) pour amener les élèves à coder un déplacement répétitif. Nous avons choisi d'utiliser une plate-forme plutôt que des robots tangibles, ce qui nous permet de nous abstraire des problèmes matériels (disponibilité et mise à disposition des robots, comportement physique...) et de faciliter les expérimentations à large échelle.

Hormis les travaux qui visent à définir une typologie de compétences ou des concepts de la PI et de leur articulation, on trouve un certain nombre de travaux plus focalisés sur les activités à mettre en œuvre.



Baratè *et al.* (2017) abordent les notions d’algorithmique et de représentation de l’information en primaire à travers des activités visant à représenter des musiques par l’intermédiaire de briques Lego.

Brackmann *et al.* (2017) ont trouvé un effet significatif d’activités débranchées sur les compétences liées à la PI. Ils concluent que ce type d’activité constitue une bonne introduction à la PI mais que des activités « en ligne » sont nécessaires pour aller plus loin.

Aggarwal *et al.* (2017), quant à eux, ont comparé l’effet des modalités tangible et numérique sur l’acquisition des compétences. Ils concluent que si la modalité tangible facilite l’appropriation de la syntaxe, la modalité numérique permet de mieux développer la capacité à visualiser et à prévoir l’effet de la séquence d’instructions. Ils suggèrent donc une utilisation limitée de la modalité tangible comme introduction avant de passer au numérique.

Dans le cadre de notre travail, nous nous sommes concentrés sur la détection et la synthèse de motifs répétitifs qui est une des composantes de la capacité d’abstraction propre à la pensée informatique. Cette compétence est abordée à travers des activités tangibles qui permettent d’élaborer cette notion de motif et d’introduire son traitement par la répétition. Les activités numériques permettent de renforcer ces concepts et de les associer à la structure de répétition utilisée en programmation visuelle puis textuelle.

## **2.2. Évaluation des compétences de la pensée informatique**

Outre les questions de l’introduction des concepts de la pensée informatique et des modalités utilisables, il existe un enjeu autour de l’évaluation des compétences acquises, que ce soit dans le cadre de recherches ou en situation d’enseignement.

Brennan et Resnick (2012) articulent la PI autour de trois dimensions : Les concepts algorithmiques (boucles, parallélisme, etc.), les pratiques (développement itératif, débogage, etc.) et les perspectives (expression personnelle, connexion avec les autres, etc.). Ils proposent d’évaluer ces dimensions à travers l’analyse des productions, des entretiens basés sur les productions et des projets de développement.

Le rapport du *Stanford Research Institute* (SRI) produit par Snow *et al.* s’intéresse aux moyens d’évaluer les compétences de la PI (résolution de

problème, abstraction, etc.) dans le contexte de l'enseignement *Exploring Computer Science* (ECS) qui se déroule sur un an au lycée aux États-Unis (Snow *et al.*, 2017). À cette fin, ils proposent des patrons de conception pour la création d'évaluations pertinentes des connaissances et des pratiques. Le rapport couvre l'évaluation des unités suivantes du cours : interaction homme-machine, résolution de problème, conception web et introduction à la programmation. Ces évaluations incluent des quiz, des problèmes, de la lecture de code et le traçage de son exécution.

Grover *et al.* (2014) combinent évaluation formative et sommative dans le cadre d'un module d'introduction des concepts de la PI d'une durée de 6 semaines dans le secondaire. Les évaluations reposent sur des questionnaires à choix multiples dont la plupart comprennent des extraits de code en Scratch. Certains exercices impliquent de remettre en ordre des blocs de code ou des activités de traçage et de debugage de code.

Seiter et Foreman (2013) proposent un cadre pour l'évaluation des compétences de la PI à l'école primaire baptisé *Progression of Early Computational Thinking* (PECT). Ce cadre s'appuie sur l'évaluation de programmes en Scratch (utilisation d'instructions spécifiques) dans le cadre de patrons de conceptions classiques liés au développement d'animations ou de jeux (par ex., animation, gestion des collisions, etc.). Ces patrons sont mis en relation avec des concepts de la PI. Le cadre proposé a été évalué sur la base des programmes disponibles sur le site Web de Scratch.

Les approches basées sur l'étude du code produit par les apprenants sont intéressantes à titre exploratoire mais chronophages. L'autre mode d'évaluation majoritaire utilise des QCM éventuellement basés sur de la lecture de code. Dans le cadre de cette étude, nous avons élaboré des questionnaires amont et aval spécifiques destinés à évaluer la capacité de reconnaissance et de synthèse des motifs dans un contexte hors programmation (plutôt sur le plan cognitif). L'analyse des traces d'activité sur la plate-forme nous donne par ailleurs des indications sur l'opérationnalisation des concepts algorithmique associés.

### **3. Cadre expérimental**

#### **3.1. Participants et organisation**

L'expérimentation de 2018 a impliqué des élèves de CM1/CM2 de 16 écoles de Villeneuve d'Ascq. Vingt classes ont participé pour un total de 447 élèves. L'âge des élèves était compris entre 8 et 10 ans avec un équilibre

des genres (49 % de filles). L'expérimentation s'est déroulée sur une semaine avec 5 classes par jour (hors mercredi). Les classes sont venues à l'université pour la journée, sous forme de sortie scolaire.

Afin de prendre en charge un aussi grand nombre d'élèves, nous faisons appel aux étudiants de deuxième année de DUT informatique, eux-mêmes encadrés par des enseignants du département informatique. La séquence pédagogique et les activités sont présentées au préalable aux étudiants de façon à ce qu'ils puissent encadrer et accompagner les élèves dans leurs apprentissages.

### 3.2. Progression pédagogique

La progression pédagogique est présentée dans la figure 1. Celle-ci inclut des activités débranchées et en ligne liées à l'identification de motifs répétitifs ainsi qu'à leur expression sous forme de séquences d'instructions et de répétitions. Nous avons trois phases décrites dans la suite. Les deux premières durent 1h30 chacune, la dernière dure 2 heures. Les élèves suivent les deux premières phases le matin et la troisième l'après-midi.

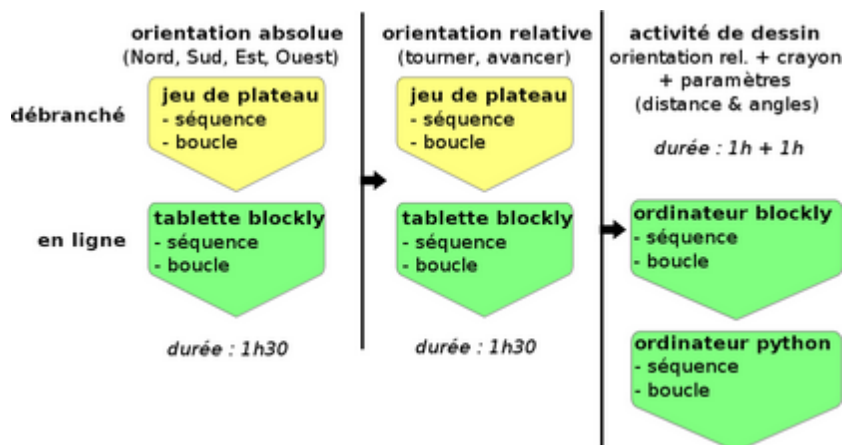
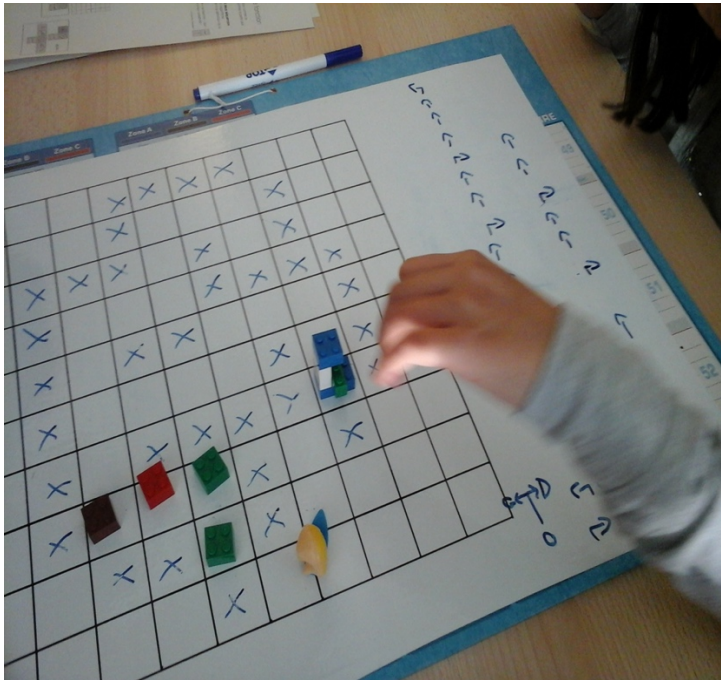


Figure 1 • Progression pédagogique

#### 3.2.1. Personnage avec orientation absolue (phase 1)

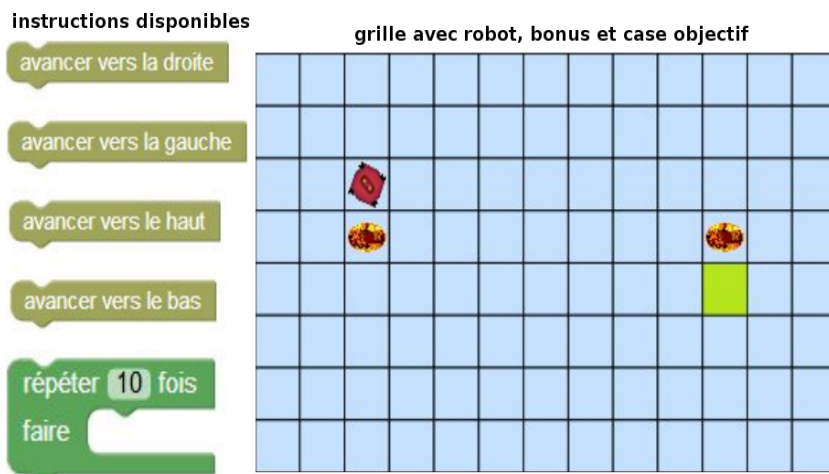
Dans cette première phase, les élèves doivent déplacer un personnage sur une grille jusqu'à une case définie en utilisant une orientation absolue (Nord, Sud, Est, Ouest). Ils démarrent l'activité avec un jeu de plateau (cf. figure 2). Les élèves travaillent en groupe de 4 à 6 et prennent à tour de rôle différentes fonctions : élaborateur d'une solution (c.-à-d. algorithme), pointeur d'instruction (qui annonce l'ordre à exécuter), et processeur (qui

exécute l'ordre en déplaçant le personnage). La complexité des chemins évolue avec l'ajout d'obstacles et de bonus. Quand la séquence d'ordres gagne en longueur, la frustration des élèves augmente. C'est à ce moment-là que l'on introduit la notion de répétition (répéter n fois).



**Figure 2 • Jeu de plateau**

Quand les concepts principaux d'instruction, de séquence, de répétition, d'exécution (et de bug) sont compris, les élèves passent sur tablette par binômes pour une activité similaire visant à renforcer les concepts et à permettre un traitement algorithmique via un langage de programmation visuelle basée sur des blocs (*Blockly*) (cf. figure 3). Pour chaque puzzle, les élèves doivent déplacer un robot jusqu'à la case destination (verte) en récupérant éventuellement des gemmes. Pour cela, les élèves disposent d'un jeu d'instructions spécifique et peuvent être limités en nombre d'instructions. Un puzzle est validé quand le robot arrive sur la case destination et a collecté toutes les gemmes. Les activités évoluent à nouveau de la séquence à la répétition pour renforcer les concepts découverts lors de l'activité débranchée.



**Figure 3 • Puzzle sur tablette avec le jeu d'instructions**

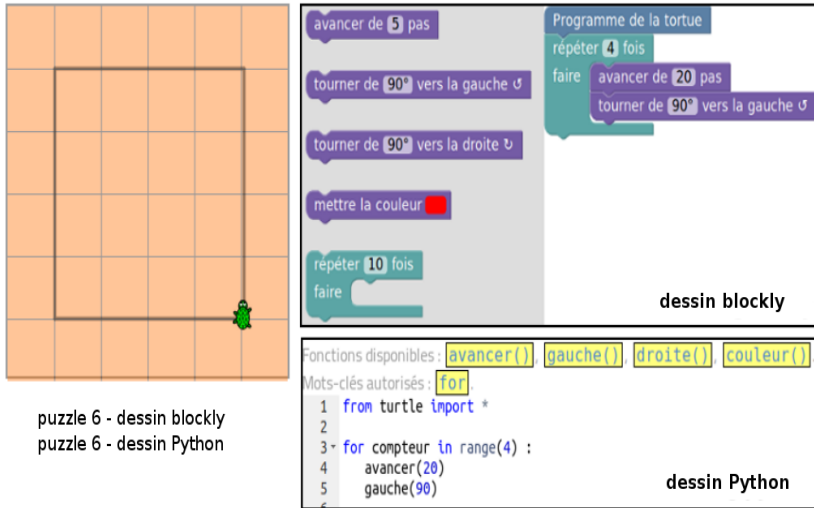
### 3.2.2. Personnage avec orientation relative (phase 2)

Cette deuxième phase suit une organisation similaire avec un premier temps sur le jeu de plateau, suivi d'un renforcement sur tablette. Le changement principal tient à l'utilisation d'un personnage orienté qui implique un jeu d'ordres différent (tourner à gauche, tourner à droite, avancer). Cela implique également de se rappeler l'orientation du personnage dans la phase de planification. Le personnage tourne obligatoirement de 90°. Cette phase prépare également à la phase de dessin suivante pour laquelle l'orientation est nécessaire.

### 3.2.3. Activité de dessin (phase 3)

La dernière phase est réalisée en salle de Travaux Pratiques avec un ordinateur par élève. Les élèves réalisent des activités de dessin avec une tortue (dans l'esprit de Logo). Le jeu d'instructions est similaire à la phase précédente avec en plus la gestion du crayon (lever et baisser pour dessiner) et la paramétrisation des instructions (par ex., avancer(distance) ou tourner à droite(angle)).

Les élèves utilisent la même plate-forme que sur les tablettes. Dans la première partie, ils continuent d'utiliser *Blockly* pour dessiner. Dans la seconde partie, nous introduisons la programmation *Python*, les faisant ainsi passer d'une programmation par blocs à une programmation textuelle. Afin de faciliter les choses, les élèves utilisent des instructions en français (par ex., avancer(10)). Toutefois, ils utilisent la syntaxe *Python* pour les répétitions.



**Figure 4 • Passage de la programmation par blocs à la programmation textuelle**

Les langages textuels sont en général introduits beaucoup plus tard dans la scolarité. Introduire le langage *Python* dans la dernière séquence sur un contexte similaire et avec la même interface (cf. figure 4) nous donne des indices sur la capacité des élèves de cette tranche d'âge à transposer leurs compétences de la programmation par blocs à la programmation textuelle.

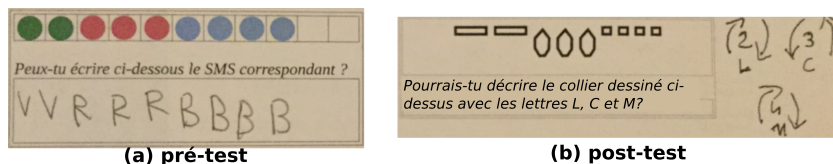
### 3.3. Collecte des données

#### 3.3.1. Pré et post tests

Les élèves passent un test au début et à la fin de la journée de manière à mesurer l'évolution de leur capacité à identifier des motifs répétitifs et à les exprimer sous la forme de répétitions.

Les tests portent sur le codage de motifs avec des lettres. Les élèves sont informés qu'ils peuvent utiliser la notation qui leur semble la plus appropriée, y compris raccourcie. Le pré test consiste à coder une représentation graphique d'une partition de musique où chaque note est représentée par une couleur (cf. figure 5a). Le post test porte sur le codage d'instructions pour la réalisation d'un collier de pâtes (cf. figure 5 b). Nous avons choisi deux contextes différents pour éviter que les élèves ne se contentent de se rappeler des motifs du pré test. Cependant, les séquences

à analyser sont strictement identiques et présentées exactement dans le même ordre.



**Figure 5 • Motifs des pré et post test**

Le tableau (cf. Tableau 1) présente les motifs utilisés, qui passent d'un corps de boucle ne contenant qu'une instruction et allant jusqu'à trois instructions pour les motifs les plus complexes. La notation présentée ici correspond au pré test mais nous avons exactement les mêmes motifs pour le post test. Le motif correspond à ce qui est présenté aux élèves et l'encodage à ce qui est attendu des élèves, la correspondance définit à quoi correspond le type de motif.

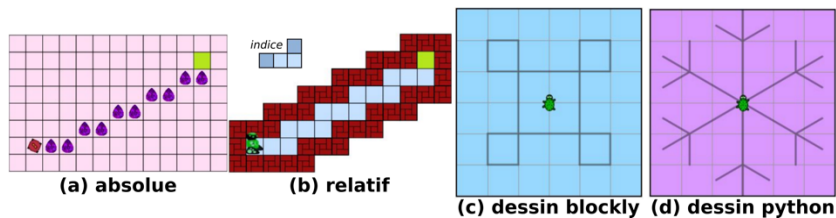
**Tableau 1 • Motifs des tests**

Type	Correspondance	Motif	Encodage
1i	motif 1 instruction	RRRRRRRRRR	11xR
Nx1i	N x motifs 1 instruction	VVRRRB BBB	2xV 3xR 4xB
2i	motif 2 instructions	BRBRBRBRBR	5x(BR)
3i+2i	motifs 2 instructions + 1 instruction	VRRVRRBBBB	2x(VRR) 4B (2x(V 2R) 4B)

### 3.3.2. Activité de programmation

Les activités de programmation ont été réalisées sur une plate-forme de l'association France-IOI. Les quatre séquences pédagogiques comprennent un ensemble de puzzles de difficulté croissante. On retrouve successivement des séquences, des répétitions à une instruction, un mélange des deux, des répétitions à plusieurs instructions, jusqu'aux boucles imbriquées. La première séquence comporte 24 puzzles avec une difficulté très progressive afin que les élèves puissent construire leurs compétences. Les séquences suivantes comprennent de 15 à 18 puzzles. Elles démarrent avec des puzzles simples (séquence), de manière à ce que les élèves puissent transposer leurs compétences sur un nouveau jeu d'instructions, puis la difficulté augmente.

La figure 6 présente quelques puzzles qui sont analysés dans la suite. La plateforme passe d'un puzzle au suivant après validation mais elle permet également de sélectionner un puzzle spécifique. Chaque séquence réalisée sur la plateforme dure 30 à 45 minutes suivant les groupes. Cela explique que les élèves n'ont pas réalisé le même nombre de puzzles suivant leur rapidité. Quand de nouveaux concepts sont introduits (par ex., la répétition), le premier puzzle comprend une aide pour la réalisation.



**Figure 6 • Puzzles difficiles dans chaque phase**

## **4. Résultats**

### **4.1. Analyses des pré et post tests**

Les tests pré et post activité ont été codés en fonction de la reconnaissance ou non des motifs par l'élève, pour arriver à une valeur entre 0 et 1. Pour les 447 élèves, nous obtenons une moyenne  $m = 0,147$ , écart-type 0,07 pour le pré test et une moyenne  $m = 0,241$ , écart-type 0,12 pour le post test.

Un t-test apparié nous permet de déterminer si la différence de moyenne est statistiquement significative. Celui-ci nous donne une valeur  $t(446) = -6,76$  ( $p < 0,0001$ ) qui confirme l'impact significatif de la séquence pédagogique sur la capacité des élèves à repérer et à encoder des motifs répétitifs.

Le Tableau 2 présente un détail de résultats pour les différents types de motifs de complexité croissante. On note que la reconnaissance des motifs à une instruction et des séquences de répétitions à une instruction augmente de manière significative (+40 % pour les deux). Le plus intéressant est probablement que la reconnaissance des motifs plus complexes (2 et 3 instructions) a progressé encore plus. Cela pourrait impliquer qu'après un apprentissage sur des motifs courts, la compétence est généralisée rapidement à des motifs plus complexes par les élèves.



**Tableau 2 • Reconnaissance des motifs (nombres d'élèves et pourcentages)**

	1i	Nx1i	2i	3i+2i
pré test	118 (27 %)	102 (23 %)	29 (8 %)	14 (4 %)
post test	166 (38 %)	140 (32 %)	70 (17 %)	56 (14 %)

La figure 5 montre le cas idéal d'un élève qui n'a utilisé aucune notation de répétition dans le pré test mais l'a appliquée avec succès dans le post test.

#### 4.2. Analyses des activités

Nous avons obtenu un export des traces d'activités des participants sur la plate-forme. Celles-ci sont malheureusement limitées car nous n'avons que la dernière validation de l'élève pour chaque puzzle et le moment où cette validation est faite. Nous n'avons donc pas l'historique des essais des élèves. La validation du puzzle peut être positive (c.-à-d. le puzzle est réussi) ou négative.

Pour chaque séquence d'activité en ligne, nous présentons un graphe indiquant pour chaque puzzle le taux de succès (nombre de validations positives/nombre de validations total) et le nombre total de validations (essais). Nous présentons également les transitions entre les niveaux de difficulté (par ex., de séquence à répétition) ce qui nous permet de visualiser les paliers de difficulté.

Pour les deux premières phases, les élèves partagent une tablette à deux et résolvent un puzzle chacun leur tour. En pratique, ils collaborent généralement pour résoudre le puzzle même s'ils n'en ont pas eu la consigne. Ceci explique le nombre d'essais maximal autour de 200. Pour la dernière phase, les élèves travaillent seuls sur ordinateur et nous avons potentiellement 447 essais (nombre d'élèves impliqués). Nous avons perdu des essais sur la première et la dernière séquence suite à des problèmes techniques, ce qui explique des chiffres inférieurs.

#### 4.2.1. Analyse des activités avec orientation absolue (phase 1)

La figure 7 montre les résultats de la première phase. Nous avons une progression mesurée de la difficulté des puzzles donnant un taux de succès supérieur à 90 %.

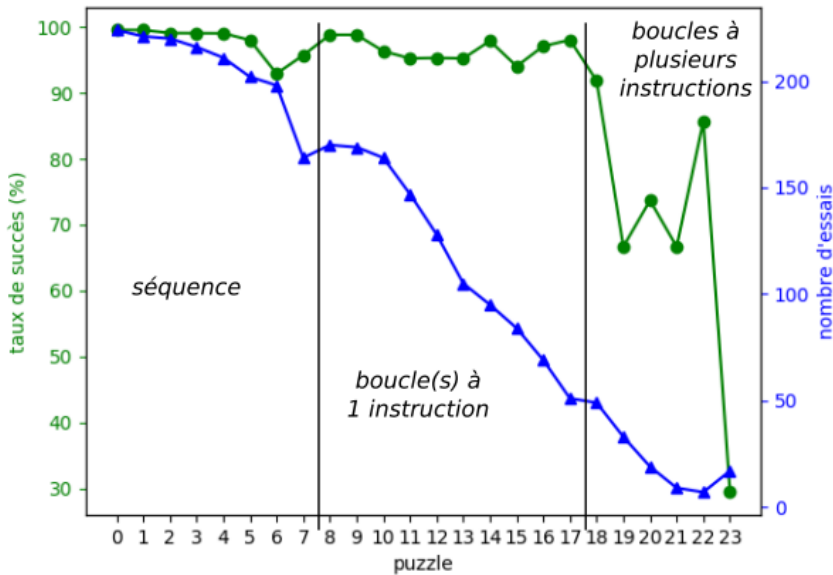
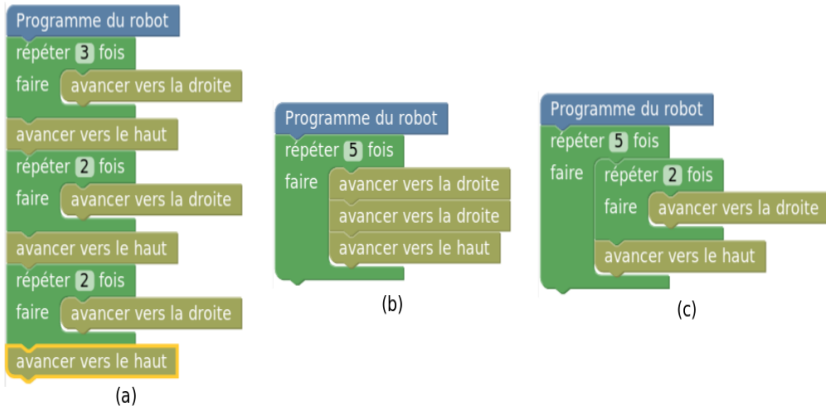


Figure 7 • Résultats de la 1<sup>re</sup> phase (orientation absolue)

On observe une diminution du nombre d'essais quand on aborde les puzzles avec répétitions, ce qui indique que certains élèves commencent à rencontrer des difficultés. Toutefois, la chute brutale correspond aux répétitions impliquant plusieurs instructions (c.-à-d. motifs plus longs) avec un taux de succès qui tombe à 66 % pour le puzzle 19 (cf. figure 6a) (le puzzle 18 est un tutoriel).

L'étude des programmes produits par les élèves permet d'observer que, dans l'ensemble, le bloc répéter est relativement bien utilisé, dans la mesure où sur la séquence on ne retrouve que trois occurrences de programme où la valeur par défaut pour la répétition a été laissée (10). Quelques élèves ont utilisé une boucle à une répétition. Cela s'est principalement produit dans les premiers puzzles comprenant une séquence d'une instruction et une répétition (puzzle 10 - 16 solutions sur 166, soit 9,6 %; puzzle 11 - 10 solutions sur 150, soit 6,6 %). Ce phénomène devient marginal dans les puzzles suivants (moins de 3 par puzzle).

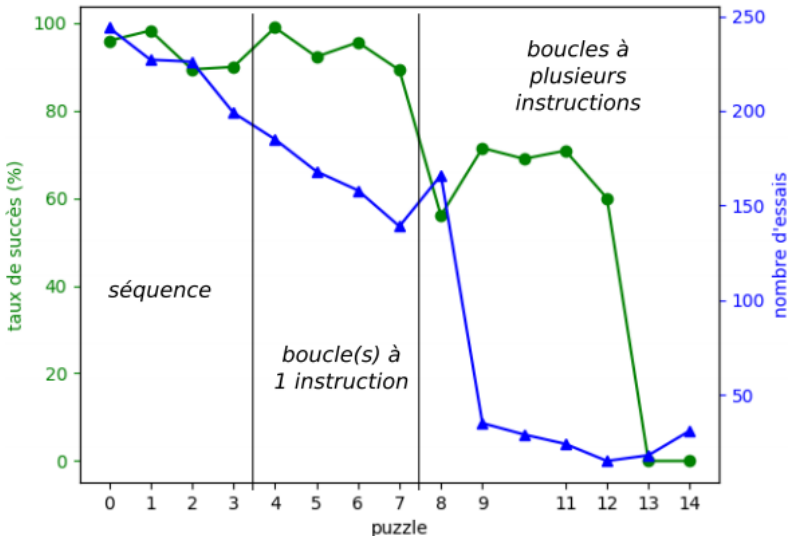


**Figure 8 • Orientation absolue : solutions pour le puzzle 19**

Si on considère plus précisément le puzzle 19, nous avons 33 solutions soumises dont 22 correctes (66 %). Parmi les solutions incorrectes, 4 n'utilisent pas de boucle (12 %) et 7 contiennent des boucles mais à une seule instruction (21 %), montrant qu'ils n'ont pas détecté le motif de longueur supérieur. La figure 8a montre un exemple de ce type de solution où l'élève a atteint la limite des blocs autorisés. Pour les solutions correctes, on a 19 solutions attendues avec une répétition de motif de longueur 3 (cf. figure 8 b) et 3 solutions avec deux boucles imbriquées (cf. figure 8c).

#### **4.2.2. Analyse des activités avec orientation relative (phase 2)**

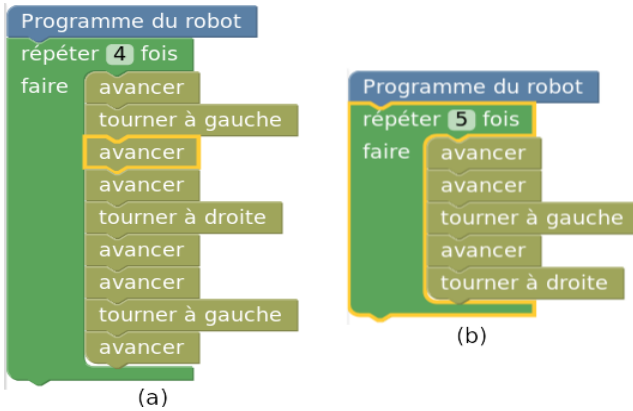
La figure 9 correspond aux résultats de la deuxième phase. Le taux de succès proche de 90 % ou supérieur indique que les élèves ont réussi à transférer leurs compétences sur le nouveau langage. On note également une amélioration de leur capacité à gérer les boucles à une instruction. En effet, on observe encore 137 essais avec un succès raisonnable pour le puzzle 7 (à comparer aux 51 essais du puzzle 17 de la phase précédente). À nouveau, passer aux répétitions impliquant plusieurs instructions présente une difficulté majeure. Le puzzle 8 (cf. figure 6 b) qui a un nombre d'essais significatif atteint un taux de succès de 56 % seulement. Et ce, malgré le fait qu'étant le premier puzzle de ce type, il y avait une indication sur le motif à traiter.



**Figure 9 • Résultats de la 2<sup>e</sup> phase (orientation relative)**

Quand on analyse les programmes produits par les élèves, on ne trouve plus qu'un nombre marginal de répétitions avec une seule itération (3 maximum sur un des puzzles), ce qui confirme la bonne maîtrise du paramétrage du bloc et un bon dénombrement des actions à réaliser.

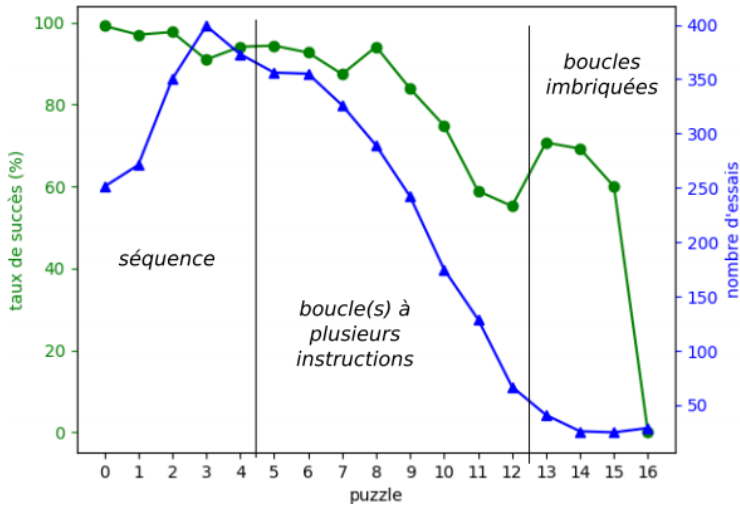
L'étude des solutions pour le puzzle 8 (cf. figure 6 b) montre 27 solutions (incorrectes) sans boucles (15 %) pour 148 solutions avec répétitions. Parmi ces dernières, 50 sont fausses (34 %). 17 de ces solutions combinent séquences et répétitions à une seule instruction, indiquant la difficulté à passer à un motif de longueur supérieure. Pour les autres, elles révèlent la difficulté à isoler le motif répété (cf. par exemple figure 10a). On retrouve au final 48 solutions conformes à la réponse attendue (27 %) (cf. figure 10 b).



**Figure 10 • Phase 2 (orientation relative) : solutions pour le puzzle 8**

#### 4.2.3. Analyse des activités de dessin en *Blockly* (phase 3)

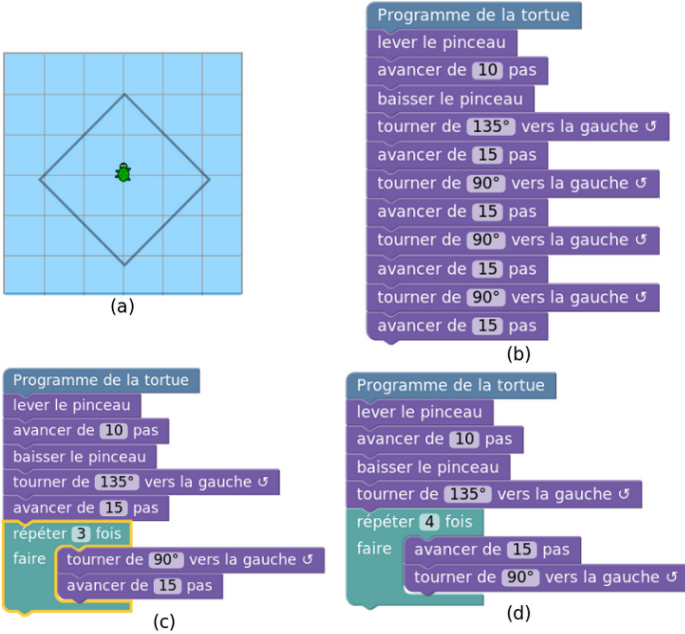
La dernière phase introduit de nouvelles difficultés. D'abord les instructions sont paramétrées. Ensuite, on commence à avoir des angles différents de  $90^\circ$  que les élèves n'ont pas encore vu en classe. À nouveau, le nombre d'essais et le taux de succès des premiers puzzles (y compris avec répétitions) indiquent que le changement de langage n'est pas un problème pour les élèves une fois que les compétences liées aux séquences et répétitions sont intégrées (cf. figure 11). Les puzzles qui combinent séquences d'instructions et répétitions comprenant plusieurs instructions restent difficiles (puzzles 9 à 11, le 9 fournissant des indices) dans la mesure où l'on voit le nombre d'essais chuter. Pour le puzzle 11 (voir figure 6c), nous avons encore 129 essais mais avec un taux de succès de 58 %. Peu d'élèves ont abordé les puzzles avec des boucles imbriquées mais avec un taux de succès légèrement supérieur à 60 %.



**Figure 11 • Résultats de la 3<sup>e</sup> phase (dessin en *Blockly*)**

Si on s'intéresse au puzzle 9 (cf. figure 12a), on a 249 essais dont 82 solutions sans boucles (32 %) - le nombre de blocs autorisés le permettait. Sur ces 82 solutions, 60 ont été validées correctes (par ex., figure 12 b). Une hypothèse pour cette proportion non négligeable de solutions basées sur la séquence est que le puzzle nécessite un angle différent de  $90^\circ$  qui présente une difficulté pour les élèves.

Parmi les 167 solutions avec boucles, 145 ont été validées par la plateforme (58 %). 79 correspondent à la solution attendue (31 %) (cf. figure 12d). Parmi les solutions non optimales, on retrouve des solutions du type de la figure 12c qui témoignent de la difficulté à bien coder le motif répétitif.



**Figure 12 • Phase 3 (dessin en *Blockly*) :  
puzzle 9 et solutions trouvées**

Le puzzle 11 (cf. figure 6c) quant à lui totalise 129 solutions soumises. On retrouve 25 solutions sans répétition qui sont toutes incorrectes (19 %). Ici, la difficulté à trouver le motif a amené une grande diversité de réponses avec 38 solutions différentes. Ces solutions ont été acceptées car la plateforme valide uniquement le tracé réalisé. Ainsi, on trouve 9 solutions (7 %) avec une répétition à 10 itérations (valeur par défaut), probablement parce que les élèves étaient concentrés sur la recherche du motif aux dépens de la réflexion sur le nombre d'itérations nécessaires. On observe également 25 solutions (19 %) utilisant une boucle imbriquée (par exemple dans la figure 13a). La solution (cf. figure 13 b) basée sur une seule répétition présente le même défaut que la solution précédente (changements de sens inutile). Enfin, la solution présentée en figure 13c représente la solution la plus courante.

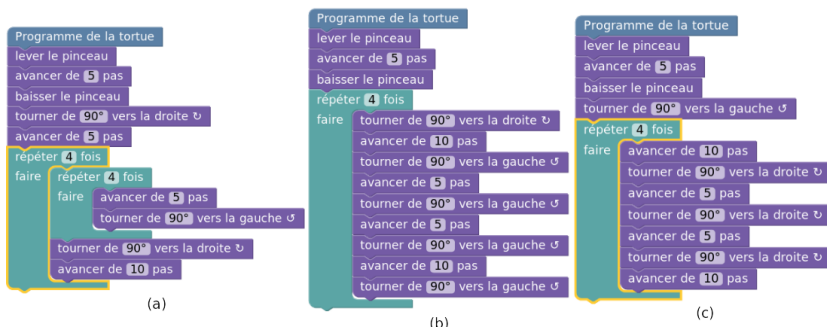


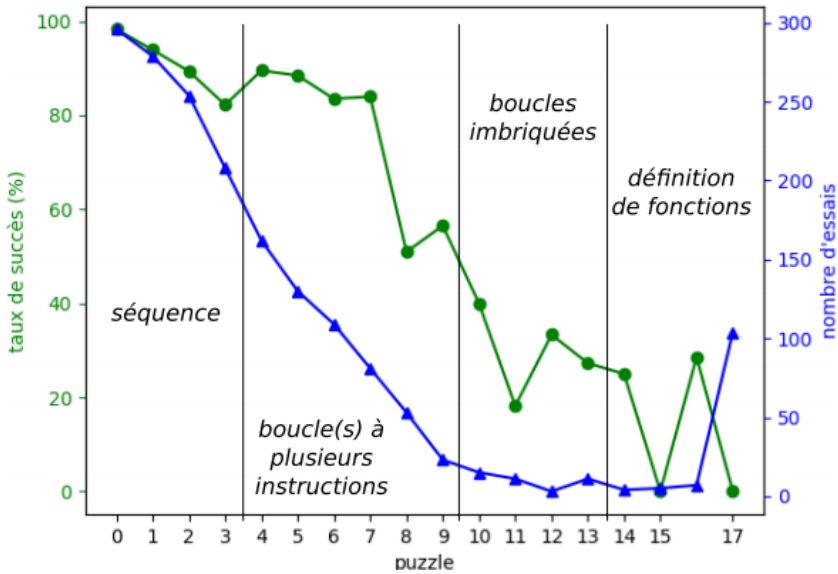
Figure 13 • Phase 3 (dessin en *Blockly*) : solutions pour le puzzle 11

#### 4.2.4. Analyse des activités de dessin en *Python* (phase 3)

La figure 4 permet de voir que les élèves peuvent utiliser des boutons pour générer les instructions afin d'éviter d'avoir à tout taper au clavier. Ils doivent malgré tout adapter les paramètres de ces instructions à leurs besoins.

Comme le montre la figure 14, la première séquence de puzzles a toujours un bon taux de succès avec 82 % et une bonne participation (nous avons perdu des traces suite à un problème technique). C'est un résultat intéressant qui montre que les élèves ont bien transposé leurs compétences d'un langage visuel à un langage textuel. Les puzzles impliquant séquence et répétition sont toujours difficiles (puzzles 8 et 9) avec un taux de succès à peine supérieur à 50 %. Les boucles imbriquées sont également un point difficile. Le puzzle 10 est un tutoriel. Le puzzle 11 présenté en figure 6d atteint seulement 18 % de taux de succès avec un faible nombre d'essais. Le dernier puzzle correspond à une activité où les élèves peuvent dessiner librement. Il n'y a donc pas de condition de validation. Le graphe montre qu'un bon nombre d'essais ont été réalisés par les élèves bien qu'on soit sur la modalité textuelle.





**Figure 14 • Résultats de la 3<sup>e</sup> phase (dessin en Python)**

L'étude des programmes produits montre les difficultés liées à la syntaxe qui apparaissent avec le langage textuel. On retrouve notamment dans les programmes non validés l'oubli du nombre d'itérations dans les répétitions ou des parenthèses enlevées par erreur dans les passages de paramètre.

## 5. Conclusion

Cet article porte sur l'apprentissage des concepts fondamentaux et des compétences de la PI par des élèves de 8-10 ans. Nous avons principalement travaillé sur la notion de motif et de traitement répétitif. Nous considérons en effet que la capacité d'identification de motifs caractéristiques et leur description synthétique constitue l'une des compétences cruciales de la PI. Pour cela, nous avons conçu une séquence pédagogique pour initier les élèves aux notions d'instruction, de séquence et de répétition. Cette séquence leur permet de pratiquer ces compétences avec différents langages (libre, *blockly* et textuel) et sous différentes modalités (débranchées et en ligne).

Cette séquence a évolué depuis nos premières expérimentations en 2014 et nous avons un nombre croissant de classes impliquées. Nous cherchons également à améliorer nos instruments de mesure pour vérifier l'impact sur les élèves (pré et post tests et analyse des activités).

Les résultats de l'expérimentation de 2018 sont encourageants. L'analyse statistique des tests montre l'impact significatif de la séquence pédagogique et nous avons une augmentation significative des élèves qui identifient les motifs répétitifs et sont capables de les synthétiser sous forme de répétitions. Il semble également qu'une fois que les élèves ont acquis cette compétence pour des motifs de longueur 2 (2 instructions dans le corps de la boucle), ils sont capables de généraliser à des motifs plus longs.

L'analyse des programmes produits montre également une bonne utilisation de la répétition dans la mesure où elle est généralement paramétrée correctement (bonne évaluation du nombre d'itérations nécessaires). La difficulté à appréhender les motifs de longueur supérieure à un se retrouve dans des programmes non optimaux où le motif est partiellement pris en compte. Comme on peut s'y attendre le passage à *Python* représente une difficulté supplémentaire avec des difficultés liées au respect de la syntaxe du langage.

Les résultats de l'analyse des activités sont à prendre avec précaution dans la mesure où, bien sûr, ils sont aidés par les étudiants voire les professeurs ou des parents. Avec un étudiant pour 4 à 6 élèves, on peut toutefois espérer que cela n'a pas un impact trop fort sur les résultats. Ce type d'activité peut raisonnablement être reproduit en classe sous forme d'ateliers avec des petits groupes d'élèves. Une fois les consignes comprises, les activités débranchées peuvent être réalisées en quasi-autonomie (activité collaborative). Les activités numériques, quant à elles, peuvent être réalisées sur tablette ou ordinateur et la plate-forme gère les rétroactions ainsi que l'avancée dans les parcours.

Les élèves reçoivent un diplôme à la fin de la journée qui donne l'adresse de la plate-forme ainsi que leur code d'identification. Nous avons observé environ 300 accès dans les jours suivants et jusqu'à deux mois après (date de la récupération des données). Toutes les séquences ont été utilisées y compris le dessin en *Python* pour laquelle il y a eu 271 essais. Ces éléments sont encourageants et montrent la motivation des élèves.

## REMERCIEMENTS

Nous remercions l'association France-IOI pour la mise à disposition de la plate-forme. Ce travail est partiellement soutenu par les projets Interreg Dig-e-Lab (<https://dig-e-lab.eu/>) et Teach Transition (<https://teachtransition.eu>).

## RÉFÉRENCES

Aggarwal, A., Gardner-McCune, C. et Touretzky, D.S. (2017). Evaluating the effect of using physical manipulatives to foster computational thinking in elementary school. Dans M. E Caspersen (dir.), *Proceedings of the 48<sup>th</sup> ACM SIGCSE Technical Symposium on Computer Science Education (SIGCSE 2017)* (p. 9-14). ACM. <https://doi.org/10.1145/3017680.3017791>

Baratè, A., Ludovico, L. A. et Malchiodi, D. (2017). Fostering computational thinking in primary school through a LEGO<sup>®</sup>-based music notation. *Procedia Computer Science*, 112, 1334-1344. <https://doi.org/10.1016/j.procs.2017.08.018>

Brackmann, C. P., Román-González, M., Robles, G., Moreno-León, J., Casali, A. et Barone, D. (2017). Development of computational thinking skills through unplugged activities in primary school. Dans E. Barendsen, P. Hubwieser (dir.), *Proceedings of the 12<sup>th</sup> Workshop on Primary and Secondary Computing Education (WiPSCE 2017)* (p. 65-72). ACM. <https://doi.org/10.1145/3137065.3137069>

Brennan, K. et Resnick, M. (2012). New frameworks for studying and assessing the development of computational thinking. Dans *Annual American Educational Research Association meeting (AERA 2012)*. [https://web.media.mit.edu/~kbrennan/files/Brennan\\_Resnick\\_AERA2012\\_CT.pdf](https://web.media.mit.edu/~kbrennan/files/Brennan_Resnick_AERA2012_CT.pdf)

Ching, Y.-H., Hsu, Y.-C. et Baldwin, S. (2018). Developing computational thinking with educational technologies for young learners. *TechTrends*, 62(6), 563-573. <https://doi.org/10.1007/s11528-018-0292-7>

Drot-Delange, B., Pellet, J. P., Delmas-Rigoutsos, Y. et Bruillard, É. (2019). Pensée informatique : points de vue contrastés. *Sticef*, 26(1), 39-61.

Gouws, L., Bradshaw, K. et Wentworth, P. (2013). Computational thinking in educational activities. Dans J. Carter (dir.), *Proceedings of the 18<sup>th</sup> ACM conference on Innovation and technology in computer science education (ITiCSE 2013)* (p. 10-15). ACM.

Grover, S., Cooper, S. et Pea, R. (2014). Assessing computational learning in K-12. Dans *Proceedings of the 2014 conference on Innovation and technology in computer science education (ITiCSE 2014)* (p. 57-62). ACM. <https://doi.org/10.1145/2591708.2591713>

Knochel, A. D. et Patton, R. M. (2015). If art education then critical digital making. Computational thinking and creative code. *Studies in Art Education*, 57(1), 21-38. <https://doi.org/10.1080/00393541.2015.11666280>

Papert, S. (1972). Teaching children thinking. *Programmed Learning and Educational Technology*, 9(5), 245-255. <https://doi.org/10.1080/1355800720090503>

Rich, K. M., Strickland, C., Binkowski, T. A., Moran, C. et Franklin, D. (2017). K-8 learning trajectories derived from research literature: Sequence, repetition, conditionals. Dans J. Tenenbergh (dir.), *Proceedings of the 2017 ACM Conference on International Computing Education Research (ICER 2017)* (p.182-190). ACM. <https://doi.org/10.1145/3105726.3106166>

Sabitzer, B., Demarle-Meusel, H. et Jarnig, M. (2018). Computational thinking through modeling in language lessons. Dans *Proceedings of the 2018 IEEE Global Engineering Education Conference (EDUCON)* (p.1913-1919). IEEE. <https://doi.org/10.1109/EDUCON.2018.8363469>

Seiter, L. et Foreman, B. (2013). Modeling the learning progressions of computational thinking of primary grade students. Dans *Proceedings of the 9<sup>th</sup> Annual International ACM Conference on International Computing Education Research (ICER 2013)*(p. 59-66). ACM. <https://doi.org/10.1145/2493394.2493403>

Snow, E., Tate, C., Rutstein, D. et Bienkowski, M. (2017). Assessment design patterns for computational thinking practices in exploring computer science. SRI International. <https://pact.sri.com/downloads/AssessmentDesignPatternsforComputationalThinking%20PracticesinECS.pdf>

Weintrop, D., Beheshti, E., Horn, M., Orton, K., Jona, K., Trouille, L. et Wilensky, U. (2016). Defining computational thinking for mathematics and science classrooms. *Journal of Science Education and Technology*, 25(1), 127-147. <https://doi.org/10.1007/s10956-015-9581-5>

Wing, J. M. (2006). Computational thinking. *Communications of the ACM*, 49(3), 33-35. <https://doi.org/10.1145/1118178.1118215>



# Liens entre performance, assiduité et questions posées et/ou questions votées en ligne dans le cadre d'une classe inversée

► **Fatima HARRAK, François BOUCHET, Vanda LUENGO**  
(LIP6, Sorbonne Université)

---

---

■ **RÉSUMÉ** • Les questions des élèves sont utiles pour leur apprentissage et pour l'adaptation pédagogique des enseignants. Nous étudions ici la nature des questions posées en ligne par les étudiants et comment le vote sur ces questions peut être lié à l'apprentissage. Nous avons donc développé un schéma de codage, puis conçu un annotateur automatique que nous avons appliqué à l'ensemble du corpus. Le résultat révèle que les votants réussissent mieux et assistent plus souvent au cours, mais le fait de poser des questions est associé à un apprentissage plus important.

■ **MOTS-CLÉS** • Question d'élève, vote d'élève, classe inversée.

■ **ABSTRACT** • *Students' questions are useful for their learning experience as well as to help teachers to adapt their pedagogy. We study here a corpus of questions asked online by students and how voting on these questions can be associated to learning. We have therefore developed a coding scheme of questions and built an automatic annotator to tag the whole corpus. The result reveals the voters perform better and attend class more often, and asking more questions is associated with better learning.*

■ **KEYWORDS** • *Student's question, student's vote, blended learning.*

## **1. Introduction**

Les questions des élèves jouent un rôle important dans le processus d'apprentissage, non seulement pour aider les élèves à mieux apprendre (Sullins *et al.*, 2015), mais aussi pour aider l'enseignant à déterminer ce qui a été compris (ou non) et à adapter sa pédagogie en conséquence. Les environnements en ligne et autres environnements informatiques pour l'apprentissage humain (EIAH) peuvent éliminer de nombreux obstacles qui empêchent les élèves de poser des questions en classe (Otero et Graesser, 2001). Nous nous intéressons ici à une formation hybride dans laquelle les étudiants doivent poser chaque semaine des questions à partir de supports de cours étudiés à distance avant le cours, pour aider les enseignants à préparer leurs séances de questions-réponses en présentiel. Cependant, compte tenu du volume de questions posées, les enseignants n'ont souvent pas assez de temps pour répondre à chaque question et doivent donc sélectionner celles auxquelles ils vont répondre. Pour les aider dans ce choix et limiter le nombre de questions, ils encouragent les étudiants à voter sur les questions déjà posées avant d'en poser de nouvelles. D'un point de vue pédagogique, cela suppose que les étudiants lisent les questions des autres, ce qui peut également avoir un impact positif en les forçant à s'interroger sur leur propre compréhension des points abordés par leurs camarades. Mais on peut aussi penser qu'un vote n'est pas exactement équivalent à une question. En effet, dans le cadre théorique Interactive-Constructive-Active-Passive (ICAP) proposé par Chi et Wylie (2014), le fait de poser une question est une activité de nature « constructive » (mobilisant des processus cognitifs tels que la recherche de lacunes dans ses connaissances et la restructuration de celles-ci), prémisses d'une activité « interactive » lorsque cette question recevra une réponse, potentiellement dans le cadre d'un échange dialogique. Ces types d'activités sont plus à même d'aider l'apprentissage qu'une activité « active » comme le fait de voter (mettant uniquement en jeu une recherche dans ses connaissances pour savoir si on saurait ou non répondre à cette question), qui elle-même est préférable à un engagement « passif » où l'on se contente de lire les questions des autres. D'après ce cadre théorique, il pourrait donc être contre-productif d'encourager un étudiant à voter plutôt qu'à poser une question. Face à cette contradiction sur la valeur d'un vote, nous avons conduit des analyses afin d'explorer comment les votes sont associés à la performance, à l'engagement des étudiants et aux questions qu'ils posent.

Plus précisément, notre objectif était de répondre aux trois questions de recherche suivantes :

**(QR1)** Quel est le lien entre le vote et la performance d'un étudiant, notamment en comparant les performances des étudiants votants qui posent des questions et de ceux qui n'en posent pas ?

**(QR2)** Le vote est-il lié à l'engagement de l'apprenant en classe et globalement ?

**(QR3)** Les étudiants votent-ils sur des questions dont la nature est différente de celle des questions qu'ils posent ?

Pour traiter ces questions de recherche, nous avons défini un schéma de codage adapté aux questions des étudiants et un système d'annotation automatique pour annoter l'ensemble du corpus de questions dont nous disposons. Dans la suite de cet article, nous proposons dans la section 2 un état de l'art introduisant plus en détail le cadre ICAP sur lequel s'appuie ce travail, comparant les différentes taxonomies de questions d'élèves existantes, et s'intéressant à l'utilité possible des questions ou des votes sur des questions dans un contexte éducatif. Nous présentons dans la section 3 le contexte de l'étude et les données utilisées, avant de décrire dans la section 4 la méthodologie de catégorisation de questions et d'annotation automatique. Enfin, nous présentons dans la section 5 les résultats des analyses effectuées pour répondre aux trois questions de recherche ci-dessus, et nous concluons avec quelques perspectives et limites de ce travail en section 6.

## **2. État de l'art**

### **2.1. Cadre théorique : ICAP**

Un aspect essentiel de notre travail est relatif à la distinction entre la valeur d'apprentissage intrinsèque associée au fait de poser une question, par opposition au fait de simplement voter sur une question déjà posée par un tiers. Dans ce cadre, les travaux de Chi et Wylie (2014) fournissent un cadre théorique particulièrement pertinent. En effet, dans sa théorie ICAP, Chi différencie 4 types d'activités : « interactive » (I), « constructive » (C), « active » (A) et « passive » (P). Dans les *activités passives*, l'apprenant se contente de recevoir le savoir sans comportement visible attestant d'un travail d'intégration des nouvelles connaissances, ce qui est souvent lié à un apprentissage en surface. Les *activités actives* sont celles qui attirent l'attention de l'apprenant (impliquant souvent un mouvement physique),

telles que regarder ou manipuler certains aspects du matériel d'apprentissage, répéter, voter, etc. Les *activités constructives* sont celles qui demandent aux apprenants d'aller au-delà de ce qui était explicitement présenté dans les supports d'apprentissage, qui peuvent contenir de nouvelles idées, telles que s'auto-expliquer, induire de nouvelles hypothèses, poser des questions, réfléchir, etc. Enfin les *activités interactives* se focalisent sur le dialogue en deux types, soit avec des experts (dialogues d'instruction), soit avec des pairs (dialogues conjoints). Les 4 types d'activités ne sont pas mutuellement exclusifs, mais au contraire hautement inclusifs : ainsi être interactif subsume être constructif (e.g. pour poser une question susceptible d'entraîner un dialogue d'instruction, il faut déjà avoir fait un travail de synthèse), et être constructif subsume également être actif (e.g. pour faire un schéma de synthèse d'un cours, il faut déjà avoir retranscrit en partie celui-ci) qui subsume le fait d'être passif (e.g. pour retranscrire il faut écouter). Dans le cadre de cet article, ni les activités passives (pour lesquelles nous n'avons pas de traces) ni les activités interactives ne sont considérées, vu le contexte particulier de cette formation hybride qui ne favorise pas la collaboration et les échanges (les étudiants ne peuvent pas répondre aux questions posées par les autres étudiants, et les enseignants ne peuvent pas répondre à toutes les questions pendant les séances dédiées aux questions-réponses).

## **2.2. Typologies de questions**

Les chercheurs ont étudié le comportement de questionnement des élèves dans divers contextes éducatifs, tels que la classe (Chin et Kayalvizhi, 2002), le tutorat (Graesser et Person, 1994) et les environnements d'apprentissage en ligne (Li *et al.*, 2014). En particulier, plusieurs taxonomies ou schémas de codage présentant différents degrés de granularité ont ainsi été proposés. Scardamalia et Bereiter (1992) se concentrent sur la distinction entre les questions fondées sur le texte et celles fondées sur le savoir (ces dernières ayant un potentiel éducatif plus fort). Bien que cette distinction soit pertinente dans notre contexte, n'ayant pas accès aux transcriptions des vidéos et diapositives avec lesquelles les étudiants ont interagi avant de poser leurs questions, il était difficile d'identifier ceci automatiquement.

D'autres chercheurs ont proposé une typologie de questions distinguant celles pouvant faire l'objet d'une investigation scientifique (par ex. : comparaison, cause à effet, prédiction, exploration) des autres (Chin et Kayalvizhi, 2002). Bien que notre but ne soit pas d'encourager à



poser un certain type de questions, cette distinction pourrait s'appliquer à notre travail, mais est difficile à réaliser sans experts du domaine. Graesser et Person (1994) ont pour leur part élaboré une taxonomie de questions posées pendant les séances de tutorat, utilisée pour la génération automatique de questions. Bien que leur taxonomie puisse être pertinente ici, certaines catégories comprenaient des « questions de raisonnement approfondi » de haute qualité, associées à des modèles de raisonnement difficiles à identifier automatiquement. Enfin, des recherches récentes (Supraja *et al.*, 2017) ont utilisé une version réduite de la taxonomie de Bloom (Bloom *et al.*, 1956) pour établir un lien entre rétroaction pratique et performance de l'apprenant en matière d'évaluation. Cette taxonomie, en raison de son origine, tend toutefois à être plus appropriée aux questions de l'enseignant qu'à celles des élèves.

### **2.3. Utilité pédagogique des questions d'élèves**

L'analyse des questions d'apprentissage a été utilisée à des fins très diverses afin d'améliorer l'efficacité de l'enseignement et l'apprentissage des élèves. Ainsi, Harper *et al.* (2003) ont étudié la relation entre les types de questions posées par les élèves de collège en physique et les notions qu'ils avaient comprises dans différents sujets. L'un des aspects de la réussite scolaire d'un élève est la compréhension conceptuelle du contenu de la matière. Les chercheurs ont trouvé qu'il n'y avait pas de corrélation significative entre le nombre de questions posées et la réussite. Toutefois, les élèves qui ont posé des questions de haut niveau ont obtenu de meilleurs résultats au test de performance conceptuelle que ceux qui n'ont posé que des questions simples, ce qui indique une relation directe entre la profondeur des questions et les connaissances conceptuelles antérieures. Graesser et Person (1994) ont également trouvé que la réussite est positivement corrélée à la qualité des questions posées par les élèves qui ont acquis une certaine expérience en tutorat, tandis que la fréquence des questions n'a pas été corrélée à la réussite. Les élèves ont partiellement autorégulé leur apprentissage en identifiant les déficits de connaissances et les comblent en posant des questions, mais ils ont besoin de formation et d'entraînement pour améliorer ces compétences.

Chin et Brown (2002) se sont focalisés sur la relation entre les questions des élèves, la nature de leur réflexion et les actions adoptées durant le processus de construction des connaissances en classe. Ils ont montré que les types de questions posées par les élèves dépendent de la façon dont ils abordent leurs tâches d'apprentissage. En effet, les questions des élèves qui

portaient sur des faits et des procédures (et qui sont typiques d'une approche d'apprentissage superficielle) ont suscité peu de discussions productives. En revanche, les questions axées sur la compréhension, la prédiction, la détection des anomalies, l'application et la planification (et qui caractérisent une approche d'apprentissage approfondie) ont amené les élèves à s'engager dans des idées de réflexion et des discussions de groupe. Ces résultats montrent que les questions posées par les élèves peuvent également refléter leur engagement, cependant poser des questions « superficielles » est généralement peu utile.

Teixeira-Dias *et al.* (2005) ont exploré les questions formulées par les élèves au cours de l'élaboration des projets de groupe pour analyser leurs comportements au lieu de leur compréhension. Les auteurs ont trouvé que les questions avaient contribué à l'engagement des étudiants en chimie, permettant d'accroître l'interaction entre l'enseignant et les étudiants et leur confiance en eux-mêmes pour formuler des questions. Par conséquent, la qualité de l'interaction en classe pendant l'apprentissage et l'enseignement de la chimie a été améliorée.

#### **2.4. Utilité pédagogique des votes d'élèves**

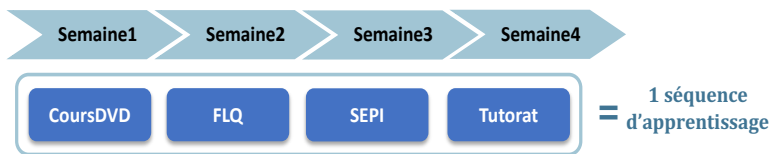
Si l'on s'intéresse maintenant aux travaux centrés sur la valeur potentielle du vote sur une question, on peut voir que les votes des élèves ont notamment été étudiés pour analyser le comportement des élèves dans les forums en ligne. Bihani *et al.* (2018) ont utilisé le nombre de votes sur les questions et réponses des étudiants et sur les réponses de l'enseignant pour révéler les paires de questions/réponses pertinentes pour les futurs cours. Zeng *et al.* (2017) ont également utilisé le nombre de votes pour détecter les messages exprimant un sentiment comme la confusion dans les forums de discussions. Ils ont constaté que les messages exprimant la confusion reçoivent un nombre important de votes. Jiang *et al.* (2015) ont analysé les étudiants considérés comme des « influenceurs » (utilisateurs dont les messages génèrent beaucoup de réponses dans les forums d'un MOOC). Ces influenceurs ont des résultats plus faibles et reçoivent moins de votes que les utilisateurs actifs (ceux qui postent régulièrement sur le forum). De même, Wong *et al.* (2015) ont analysé les votes (positifs et négatifs) sur les messages et les commentaires des utilisateurs actifs. Contrairement à Jiang *et al.* (2015), ils ont constaté que les utilisateurs actifs sont aussi des utilisateurs influents qui apportent généralement une contribution positive aux discussions du forum du MOOC. Les votes des élèves ont donc

surtout été utilisés pour analyser le comportement des élèves, mais la nature des questions votées n'a apparemment pas encore été explorée.

Dans l'ensemble, les typologies de questions proposées jusqu'à présent dépendent essentiellement du contexte, et nous avons décidé de définir un nouveau schéma de codage utilisant une approche fondée sur les données. Dans cet article, nous nous intéressons à l'analyse de la nature des questions posées et des questions votées par les étudiants.

### 3. Contexte et données de l'étude

Nous avons considéré l'ensemble des questions posées par des étudiants de 1<sup>re</sup> année de médecine et pharmacie d'une université française en 2012-2013. 1608 étudiants étaient inscrits cette année-là, une partie d'entre eux seulement ayant posé des questions. La Faculté de médecine dispose d'un système de formation hybride pour ses étudiants de 1<sup>re</sup> année (PACES). Chaque semestre se termine par un concours (en janvier et mai) éliminatoire (seule une partie des étudiants qui ont échoué au concours est autorisée à repasser l'année une seule fois, les autres devant se réorienter). Chaque unité d'enseignement est composée de deux à six séquences de 4 semaines (cf. figure 1).



**Figure 1 • Les quatre activités d'une séquence d'apprentissage sur quatre semaines**

Dans chaque séquence, la première semaine consiste à étudier le cours sur DVD-ROM ou sur le site Medatice (diapositives + vidéo du professeur). La deuxième semaine est consacrée à la Formulation en ligne des questions (FLQ) pour les enseignants : ces questions concernent exclusivement les cours multimédias étudiés la semaine précédente. Les élèves peuvent voir les questions posées uniquement par les élèves de leur groupe (environ 200 élèves par groupe) et voter pour celles auxquelles ils veulent aussi une réponse, mais il ne leur est pas possible de les commenter ou d'y répondre. En fin de semaine, les questions sont envoyées par courriel aux enseignants intervenant la troisième semaine, qui les utilisent pour structurer leurs

sessions d'enseignement interactives en classe (SEPI). Au cours de ces sessions, l'enseignant répond à certaines questions posées en ligne par les étudiants. La quatrième semaine est consacrée à des séances de tutorat afin de tester les connaissances acquises lors de la séquence de formation par le biais d'un autotest utilisant des questions à choix multiples (QCM), qui sont ensuite corrigées par un professeur auxiliaire. Il y a deux séances de tutorat de 2 heures par semaine. Chaque étudiant peut vérifier individuellement ses notes et son classement par rapport à l'ensemble de la promotion, et il est nécessaire de s'être connecté à la plateforme de questions pour pouvoir consulter ses notes.

Pour chacun des 13 cours, nous avons donc 2 à 6 ensembles de questions (un par séquence) posées au total par 429 étudiants (6457 questions au total) et votées par 672 étudiants (10 951 votes) pendant la deuxième semaine de chaque séquence. La répartition des questions est inégale (cf. Tableau 1), avec plus de questions au 1er semestre, notamment car certains étudiants sont obligés d'arrêter à la fin de celui-ci, ce qui explique qu'il y ait moins d'étudiants au 2<sup>e</sup> semestre. On note que seul un élève sur quatre a posé au moins une question, ce qui peut être lié à l'encouragement à voter au lieu de poser des questions (pour forcer à lire les questions des autres mais aussi pour réduire le nombre de questions reçues par courriel).

**Tableau 1 • Distribution des questions posées par cours**

BCH	BPH	HBD	BCE	ANT	PHS	SSH	ICM	MAT	Spec.
19 %	17 %	15 %	11 %	10 %	9 %	8 %	6 %	3 %	1 %

BCH = Biochimie, BPH = Biophysique, HBD = Histoire et biologie du développement, BCE = Biologie cellulaire, ANT = Anatomie, PHS = Physiologie, SSH = Santé, société, humanité, ICM = Initiation à la connaissance du médicament, MAT = Mathématique, Spécialité = Pharmacie, Odontologie, Maïeutique

#### **4. Méthode de catégorisation et d'annotation des questions**

Comme vu dans l'état de l'art, les typologies de questions proposées dépendent principalement du contexte étudié et fournissent rarement un ensemble complet de mots-clés pour permettre une identification automatique de questions, et encore moins des outils dédiés permettant cette classification. Notre objectif est de fournir des catégories de questions qui prennent en compte l'intention de l'élève. Par conséquent, nous avons décidé de définir notre propre schéma de codage pour identifier le type des questions posées par les étudiants, en utilisant une approche ascendante fondée sur les données. Nous présentons ici la démarche suivie pour construire l'annotateur associé.

#### 4.1. Méthodologie de catégorisation

Afin d'identifier la nature des questions posées par les étudiants, nous avons travaillé sur un échantillon de 800 questions (12 % du corpus) issues de deux cours (BCH et HBD), considérés par l'équipe pédagogique comme étant parmi les plus difficiles et ayant suscité le plus de questions (cf. Tableau 1). Cet échantillon a été divisé en 4 sous-échantillons de 200 questions pour appliquer 4 étapes successives de catégorisation.

(1) L'**étape de découverte** consistait à regrouper empiriquement des phrases ayant des similitudes pour en extraire des concepts significatifs. Bien que les enseignants demandent aux étudiants de poser des questions simples (c.-à-d. d'éviter des questions comme « Pourriez-vous expliquer à nouveau X ? De plus, Y n'était pas clair »), 40 % des questions pouvaient être divisées en plusieurs questions indépendantes. Une fois les phrases segmentées en questions dites simples, nous avons regroupé celles dont la structure (par ex. « qu'est-ce que X ? » et « qu'est-ce que Y ? ») et la sémantique (par ex. « qu'est-ce que X ? » et « pourriez-vous définir X ? ») semblent similaires. Des groupes de questions ont ensuite reçu des *étiquettes* (par ex. « définition d'un concept »), chaque étiquette est associée à un groupe de questions. Puis nous avons identifié les exclusions mutuelles entre étiquettes (par ex. une question simple ne peut pas être à la fois une vérification et une demande de réexplication). Cela nous a conduits à définir le concept de « dimension », ensemble d'étiquettes de type de questions similaires mais mutuellement exclusives (par ex. une question ne peut pas être à la fois une « vérification » et une « ré-explication » au sein de la première dimension car vérifier suppose de proposer soi-même une réexplication en premier lieu - bien sûr il peut cependant y avoir un enchaînement des deux demandes dans 2 propositions différentes). Chacune de ces étiquettes individuelles (« vérification », « réexplication »...) sont des valeurs pouvant être associées à une dimension. Chaque question simple peut alors être associée à une annotation dans ce schéma de codage en choisissant une valeur par dimension.

(2) L'**étape de consolidation** consistait à annoter le deuxième sous-échantillon pour valider les dimensions et les valeurs précédemment identifiées. Cela a conduit à divers ajustements des dimensions pour s'assurer qu'elles étaient bien indépendantes les unes des autres (par exemple l'ajout de la valeur « correction » dans la dimension Dim2, non identifiée précédemment). Parallèlement, les dimensions identifiées ont

été revues et validées par un professeur expert enseignant dans le cadre de PACES, qui a estimé que les catégories étaient potentiellement pertinentes pour analyser les questions des étudiants et ainsi pouvoir intervenir ensuite.

(3) Dans l'**étape de validation**, nous avons effectué deux annotations indépendantes pour valider l'ensemble de nos catégories sur le troisième sous-échantillon de 200 phrases. Deux annotateurs humains ont utilisé comme référence unique le schéma de codage créé à la fin de l'étape précédente pour annoter chaque segment (238 au total). À l'issue de l'étape précédente, trois dimensions avaient été identifiées : Dim1 (relative au type de question), Dim2 (relative à la modalité d'explication), Dim4 (facultative, annotée uniquement si la question est une vérification, relative à la nature de ce qui est vérifié). La dimension appelée « Dim3 » plus loin n'existait pas encore à cette étape. Les annotateurs humains ont fait deux annotations distinctes et indépendantes sur chaque dimension, et leur accord a été évalué à l'aide du Kappa de Cohen (Arstein et Poesio, 2008). Le Kappa est un score d'accord entre -1 et 1, où 1 correspond à un accord parfait et 0 à un accord uniquement explicable par le hasard (ex : en prédisant systématiquement « pile » après un lancer de pièce, bien que l'on ait raison 1 fois sur 2, cet accord entre la prédiction et la réalité s'explique uniquement par le hasard et correspondrait à un kappa de 0). Les Kappas obtenus ici sont  $K1 = 0,72$ ,  $K2 = 0,62$  (où  $K1$  et  $K2$  correspondent respectivement au Kappa de Dim1 et Dim2) soit bien au-dessus de 0 et témoignent donc d'un accord fort non uniquement explicable par le hasard. Pour Dim4, en raison de son caractère facultatif, les deux annotateurs n'ont pas nécessairement annoté les mêmes questions (annotateur 1 : 82 questions ; annotateur 2 : 68 questions) : sur les 68 en commun, le kappa valait 0,66. Puis ils se sont rencontrés pour discuter et résoudre les désaccords, ce qui a conduit à un affinement final des catégories (par exemple, séparation des catégories Dim1 et Dim4, ajout de la catégorie Dim3). Finalement, tout l'échantillon (600 phrases) a été réannoté sur les 4 dimensions pour tenir compte des changements et fournir une référence à laquelle comparer l'annotation automatique. Cette version finale du schéma de codage est présentée dans le Tableau 2. Une annotation de question peut donc être vue comme un vecteur de 4 valeurs (ex : « Pourriez-vous réexpliquer la différence entre un composé ionisable et un partiellement ionisable ? » marquée comme « Ree » sur Dim1, « Lie » sur Dim3 et aucune valeur « 0 » pour les dimensions 2 et 4, c.-à-d. [Ree,0,Lie,0]).

(4) Finalement, dans l'étape d'évaluation, le dernier sous-échantillon de 200 segments a été annoté manuellement par les deux annotateurs experts (avec un kappa accru de 0,83 sur Dim1, 0,76 sur Dim2 et 0,47 sur Dim3). Ce sous-échantillon, non utilisé pour l'entraînement de l'annotateur automatique, a été utilisé pour son test.

Le schéma de codage proposé est donc issu de l'expertise humaine puisque les 14 catégories ont été définies par les chercheurs, puis revues et validées par un enseignant expert du domaine.

**Tableau 2 • Schéma de codage créé à partir de l'annotation manuelle**

<b>Dim1</b>	<b>Type de question</b>	<b>Description</b>
Ree	Réexpliquer/redéfinir	Demander de revenir sur un concept déjà expliqué
App	Approfondir un concept	Approfondir une connaissance, clarifier une ambiguïté ou demander plus de détails pour mieux comprendre
Ver	Validation/vérification	Vérifier ou valider une hypothèse
<b>Dim2</b>	<b>Modalité d'explication</b>	<b>Description</b>
Exe	Exemple	Exemple d'application (cours/exercice)
Sch	Schéma	Schéma d'application ou explication sur ce dernier
Cor	Correction	Correction d'un exercice en cours/examen
<b>Dim3</b>	<b>Type d'explication</b>	<b>Description</b>
Def	Définir	Définir un concept ou un terme
Man	Manière (comment ?)	Demander comment procéder
Rai	Raison (pourquoi ?)	Demander la raison
Rol	Rôles (utilité ?)	Demander l'utilité/fonction
Lie	Lien entre concepts	Vérifier le lien entre deux concepts, le définir
<b>Dim4</b>	<b>Type de vérification (facultatif)</b>	<b>Description</b>
Err	Erreur/contradiction	Détecter une erreur/contradiction dans cours ou dans l'explication de l'enseignant
Con	Connaissances du cours	Vérifier une connaissance
Exa	Examen	Vérifier une connaissance attendue à l'examen

## **4.2. Annotation automatique**

Afin d'annoter l'ensemble de questions posées par les étudiants, un outil semi-automatique à base de règles et de mots clés pondérés manuellement a été utilisé dans un premier temps pour segmenter et annoter les questions automatiquement. Bien qu'efficace sur les questions qu'il annote ( $\kappa$  élevé), certaines questions ne sont pas annotées par cet outil (Harrak *et al.*, 2018) : en effet, cet outil dépend de mots-clés pondérés manuellement, et certaines dimensions dans notre schéma de codage n'ont pas de mots clés explicites pour les annoter (par ex. connaissances en cours dans la Dim4). Par conséquent, nous avons envisagé d'utiliser une annotation entièrement automatisée basée sur des techniques d'apprentissage automatique sur le corpus des questions, indépendantes de mots-clés pondérés manuellement. Les différentes étapes suivies sont décrites dans ce qui suit et sont résumées dans la figure 2.

Pour ce qui est de l'étape de segmentation préalable à l'annotation, nous utilisons un système de détection de la limite de la phrase intégré dans NLTK (Kiss et Strunk, 2006), qui est l'un des systèmes de traitement automatique de la langue fonctionnant en français. Il repose sur une approche dite non supervisée, et a été largement testé sur différentes langues et sur différents genres de textes. Il permet d'obtenir de bons résultats sans autres modifications ou ressources spécifiques à la langue. Bien que les questions de certains élèves puissent être mal rédigées et mal formulées, la méthode de segmentation semble fonctionner assez bien dans ce contexte. Il convient également de noter qu'en pratique, lors de l'annotation manuelle des segments, aucun des experts humains n'a trouvé une situation où il estimait que le segment fourni aurait dû être plus segmenté qu'il ne l'était.

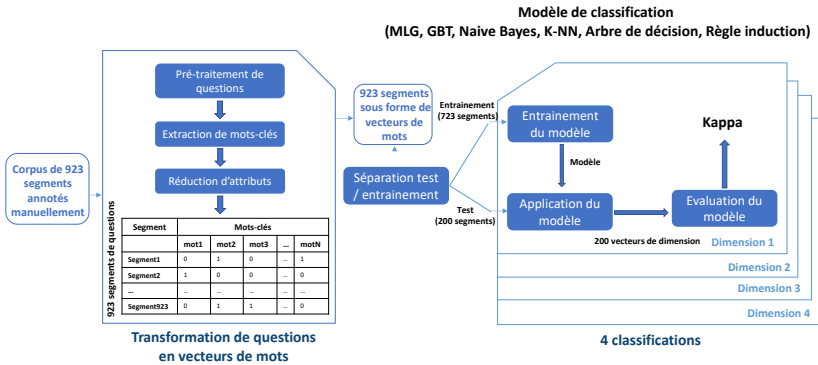
La première étape a consisté à transformer les 923 segments annotés manuellement en vecteurs de mots. Tout d'abord, nous avons utilisé la version française de WordNet (Sagot et Fišer, 2008), base de données lexicale reliant des concepts sémantiques entre eux dans une ontologie selon une variété de relations sémantiques (telles que synonymie et hyperonymie) afin de ramener différentes expressions synonymes à une même expression dans les questions. Par exemple pour la valeur « Rai » dans Dim3, les mots synonymes « cause », « raison » et « motif » sont remplacés dans le texte par « pourquoi ». L'objectif étant de diminuer la diversité lexicale et de renforcer certaines expressions pour le traitement. Nous avons effectué par la suite un ensemble de prétraitements classiques



sur le corpus de 923 segments: *tokenisation*, *racinisation*, suppression de ponctuation et de *stopwords* (mots creux non porteurs de sens), etc. Puis, nous avons extrait tous les unigrammes et bigrammes (n-grammes avec  $n = 1$  et  $n = 2$  respectivement), avec une approche de type *sac de mots*, et compté leurs occurrences dans l'échantillon de 600 questions (723 segments) de l'étape de validation. Chaque segment est représenté par un vecteur de mots (nombre d'occurrences de chaque unigramme/bigramme extrait sur chaque segment). Le nombre de n-grammes étant très important par rapport au nombre de segments, nous avons réduit celui-ci pour conserver les mots-clés les plus importants et les plus significatifs en utilisant une technique de sélection d'attributs (suppression des n-grammes les moins fréquents et corrélés).

La deuxième étape a consisté à entraîner un classifieur pour annoter automatiquement chaque valeur (ou étiquette) de dimension (par ex. « réexpliquer »). Nous avons testé 6 techniques de classification différentes telles qu'implémentées dans RapidMiner (Modèle linéaire généralisé, Gradient Boosted Trees, Arbre de décision, K-NN, Règle d'induction et Naïve Bayes, avec différentes valeurs d'hyperparamètres testées pour chacune) sur chaque dimension séparément, les dimensions étant conçues comme indépendantes. Chaque classifieur est entraîné en prenant en entrée un ensemble de vecteurs de mots représentant les 723 segments de l'ensemble d'entraînement, et l'étiquette à prédire est la valeur associée manuellement au segment dans cette dimension.

Le modèle est ensuite évalué sur un échantillon indépendant de 200 segments sans étiquettes, afin d'assurer une bonne estimation de la performance sur des données non vues. Enfin, nous avons calculé les valeurs Kappa entre les valeurs prédites par le classifieur et les valeurs correspondantes trouvées par l'annotation manuelle. Les meilleurs résultats ont été obtenus par l'algorithme Gradient Boosted Trees avec un Kappa moyen sur chaque dimension de 0,70 (cf. Tableau 3) - une valeur suffisamment élevée pour appliquer l'annotation automatique au corpus complet.



**Figure 2 • Processus d’annotation à base d’apprentissage automatique**

**Tableau 3 • Kappas obtenus entre les différentes techniques de classification utilisées et l’annotation experte de référence**

Dimension	Modèle linéaire généralisé (GLM)	Gradient Boosted Trees (GBT)	Naive Bayes	K-NN (K = 2)	Arbre décision (C4.5)	Règle induction
Dim1	0,68	<b>0,70</b>	0,29	0,57	0,36	0,70
Dim2	0,17	0,77	0,10	0,43	<b>0,79</b>	0,37
Dim3	<b>0,69</b>	0,63	0,37	0,61	0,63	0,58
Dim4	0,62	<b>0,66</b>	0,38	0,60	0,13	0,66

### 5. Analyse de différences entre les questions sur lesquelles votent les étudiants et les questions qu’ils posent eux-mêmes

À l’issue des travaux exposés dans la section précédente, nous disposons d’un annotateur automatique que nous avons pu appliquer au corpus complet de questions, ce qui a permis d’obtenir un ensemble de 6457 questions annotées automatiquement. Nous avons concentré notre analyse sur quatre cours ayant généré le plus de questions (cf. Tableau 1) et considérés par les enseignants comme les plus difficiles : BCH, HBD, BCE et ANT. Les trois premiers ont lieu au premier semestre et ANT au second semestre. Nous n’avons pas fusionné les questions des différents cours, car des études antérieures sur ces données avaient montré des différences

significatives entre les cours (Harrak *et al.*, 2019 ; Harrak *et al.*, 2018). Ceci est lié au fait que la dynamique des questions semblait être un indicateur pour distinguer les étudiants, mais celle-ci est très liée au cours. De plus, considérer les cours séparément permet de vérifier si des tendances similaires apparaissent d'un cours à l'autre.

Pour répondre aux questions de recherche, nous avons distingué 4 sous-populations sur chacun des cours considérés en fonction de l'activité des étudiants en distinguant : Q pour les étudiants « ayant posé au moins une question » (NQ sinon), et V pour les étudiants « ayant voté sur au moins une question » (NV sinon). En croisant les deux, cela donne donc les 4 sous-populations suivantes : QV, QNV, NQV et NQNV. Dans la suite de cet article, nous présentons l'analyse du lien entre vote et performance en section 5.1, la relation entre vote et engagement en section 5.2, la comparaison de la nature entre questions posées et questions votées en section 5.3 et un tableau de synthèse des résultats en section 5.4.

## **5.1. Analyse de lien entre questions, vote et performance**

Pour examiner la QRI (c.-à-d. le lien entre questions/vote et performance d'un étudiant), nous avons étudié dans un premier temps le lien entre questions et performance pour les étudiants ayant posé des questions (Q) et ceux qui n'en ont pas posées (NQ) et le lien entre vote et performance pour les votants et non votants en section 5.1.1. Ensuite, nous avons analysé de manière plus fine le lien entre le vote et la performance en section 5.1.2, notamment pour les étudiants qui posent des questions (QV vs. QNV) et ceux qui n'en posent pas (NQV vs. NQNV).

### **5.1.1. Lien entre questions et performance et lien entre vote et performance**

#### **5.1.1.1. Méthode**

Pour évaluer la performance, nous avons considéré pour chaque étudiant et sur chaque cours : (1) la note moyenne obtenue sur les QCM du cours (NotMoy, sur 20), qui peut donner une mesure de l'impact à court terme des questions posées, et (2) la note finale obtenue au concours à cette matière (NotFin, sur 20), qui peut donner une mesure de l'effet à plus long terme.

Ensuite, pour chacun des 4 cours considérés, nous avons effectué des comparaisons 2 à 2 de la performance obtenue entre d'une part les étudiants qui ont posé des questions (Q, constituée de QNV et QV) sur ce

cours et ceux qui n'en ont pas posé (NQ, constituée de NQV et NQNV), puis d'autre part la population votante (V, constituée de QV et NQV) et la population non-votante (NV, constituée de QNV et NQNV). Pour ces 2 variables (NotMoy et NotFin), nous avons utilisé des tests Mann-Whitney U (MacFarland et Yates, 2016) au lieu de t-tests (les distributions ne suivant pas une loi normale). Nous avons effectué 2 fois 7 tests (2 comparaisons de population avec 4 cours et 2 variables à chaque fois, sauf NotFin, manquante pour BCE). Nous rapportons une taille d'effet estimée, calculée comme suit :  $r^2 = \eta^2 = Z^2/n$  où  $Z$  représente le score  $z$  associé à la valeur  $p$  du test et  $n$  le nombre d'élèves de ce groupe (Fritz *et al.*, 2012). Les seuils de significativité ont été corrigés par la méthode de Holm-Šidák (Abdi, 2007) pour éviter les erreurs de type I (rejet de l'hypothèse nulle alors qu'elle est vraie). Nous avons également utilisé la correction de Yates sur ces tests pour tenir compte de la continuité lorsqu'une cellule du tableau de contingence avait un nombre inférieur à 5 et reporté l'ampleur d'effet en utilisant le  $V$  de Cramér corrigé, noté  $\tilde{V}$  (Bergsma, 2013).

### **5.1.1.2. Analyse des résultats**

Les résultats des tests pour effectuer les 2 comparaisons de population (Q vs. NQ et V vs. NV) avec 4 cours et deux variables à chaque fois (NotMoy et NotFin) sont décrits dans ce qui suit et sont résumés dans le Tableau 4. Les résultats significatifs sont mis en gras et associés à une valeur  $p < .001$  après la correction de Holm-Šidák.

Pour les étudiants ayant posé des questions (Q) et ceux qui n'en ont pas posées (NQ), il n'y avait pas de différence statistiquement significative pour les deux variables NotMoy et NotFin dans chacun des 4 cours.

En ce qui concerne les étudiants ayant voté à des questions (V) par rapport à ceux qui n'ont voté à aucune (NV), 2 résultats significatifs (cf. Tableau 4) ont été obtenus (sur 7 tests) : les étudiants qui ont voté (V) ont des notes finales plus élevées que ceux qui n'ont pas voté (NV) pour le cours ANT ( $U = 56357.5, p = .004, \eta^2 = .004$ ). La tendance est inversée pour la NotMoy pour le cours HBD ( $U = 211938.5, p < .001, \eta^2 = .010$ ). La valeur de  $p$  indique que les tests sont statistiquement significatifs, la valeur  $U$  du test est à comparer à la valeur maximale qui est le produit de la taille des 2 échantillons considérés, tandis que  $\eta^2$  indique la force de l'effet d'une variable sur l'autre.

**Tableau 4 • Comparaison des étudiants (Q vs. NQ et V vs. NV) en termes de performance**

Cours	Q vs. NQ		V vs. NV	
	Not-Moy	NotFin	NotMoy	NotFin
BCH	.477	.584	.005	.027
HBD	.408	.066	<b>.000*</b>	.015
BCE	.080	N/A	.551	N/A
ANT	.540	.020	.740	<b>.004*</b>

\* significatif avec  $p < .05$  après correction de Holm-Šidák (.000 signifie  $p < .001$ )

## 5.1.2. Liens croisés entre questions, votes et performance

### 5.1.2.1. Méthode

Afin d'analyser de manière plus fine la relation entre les votes et les questions en termes de performances, nous avons fait des comparaisons 2 à 2 entre QV et QNV (pour analyser le vote chez les étudiants ayant posé des questions) et entre NQV et NQNV (pour analyser le vote chez les étudiants qui n'ont pas posé de questions), sur chacun des 4 cours, pour les deux variables NotMoy et NotFin. Pour ces 2 variables, nous avons utilisé les mêmes tests qu'en section précédente et effectué 2 fois 7 tests (2 comparaisons de population avec 4 cours et 2 variables à chaque fois, sauf NotFin, manquante pour BCE) et les seuils de significativité ont été corrigés par la méthode de Holm-Šidák pour éviter les erreurs de type I.

### 5.1.2.2. Analyse des résultats

Les résultats des tests de comparaison des votants (QV vs. QNV et NQV vs. NQNV) en termes de performance (note moyenne et note finale) et statistiques descriptives (quartiles et médiane) des 4 sous-populations sur chacun des 4 cours sont présentés dans les Tableaux 5 et 6.

En ce qui concerne QV par rapport à QNV, 1 seul résultat significatif a été obtenu sur les cours du premier semestre (BCH, HBD, BCE) : pour HBD, parmi les étudiants ayant posé des questions, ceux n'en ayant pas voté ont eu une meilleure note finale que les étudiants qui en ont voté ( $U = 2977,5$ ,  $p = .002$ ,  $\eta^2 = .040$ ). En revanche, pour ANT, en dépit d'effectifs plus réduits, les étudiants qui ont posé des questions et voté ont mieux réussi que ceux n'ayant fait que poser des questions, aussi bien aux QCM du cours ( $U = 1452$ ,  $p < .001$ ,  $\eta^2 = .127$ ) qu'au concours final ( $U = 1494,5$ ,  $p < .001$ ,  $\eta^2 = .155$ ).

En ce qui concerne les NQV par rapport à NQNV, 3 résultats significatifs ont également été obtenus (sur 7 tests) : pour BCH ( $U = 112\,024$ ,  $p = 0,001$ ,  $\eta^2 = 0,006$ ) et HBD ( $U = 121258,5$ ,  $p < .001$ ,  $\eta^2 = .016$ ), les étudiants n'ayant pas posé de questions et n'ayant pas non plus voté ont eu une meilleure note sur les QCM que ceux ayant uniquement voté. Ce résultat se retrouve également au niveau du concours final pour HBD ( $U = 129\,974$ ,  $p < .001$ ,  $\eta^2 = .007$ ). Aucune différence n'a été observée au cours du deuxième semestre (ANT). En résumé, lorsqu'une différence a été observée, les élèves qui ont voté (sans poser de question) ont obtenu des résultats inférieurs, tant dans le cours que dans l'ensemble.

**Tableau 5 • Comparaison des votants en termes de Note Moyenne (NotMoy) pour les étudiants qui posent des questions (QV vs. QNV) et ceux qui n'en posent pas (NQV vs. NQNV)**

	QV vs. QNV	NQV vs. NQNV	QV				QNV				NQV				NQNV			
	p	P	N	Q1	Md	Q3	N	Q1	Md	Q3	N	Q1	Md	Q3	N	Q1	Md	Q3
BCH	.049	.001*	181	6,5	8,5	11,5	59	7,7	9,7	12,5	217	5,6	8,0	10,1	980	5,5	8,0	10,9
HBD	.176	.000*	154	7,7	10,1	13,8	53	8,2	11,7	14,0	252	6,0	8,7	11,2	956	6,0	9,2	12,5
BCE	.128	.039	83	7,4	10,4	12,6	47	6,4	9,0	11,7	117	6,0	8,0	10,8	1133	5,0	7,7	10,8
ANT	.001*	.005	42	11,2	13,8	15,3	46	6,4	10,2	13,3	23	8,6	10,0	12,7	968	5,0	8,0	11,8

\* significatif avec  $p < .05$  après correction de Holm-Šidák (.000 signifie  $p < .001$ )

**Tableau 6 • Comparaison des votants en termes de la Note Finale (NotFin) pour les étudiants qui posent des questions (QV vs. QNV) et ceux qui n'en posent pas (NQV vs. NQNV)**

	QV vs. QNV	NQV vs. NQNV	QV				QNV				NQV				NQNV			
	p	P	N	Q1	Md	Q3	N	Q1	Md	Q3	N	Q1	Md	Q3	N	Q1	Md	Q3
BCH	.080	.015	168	5,2	8,2	11,5	56	6,3	10,3	13,3	197	4,2	6,5	9,7	880	3,7	6,7	10,5
HBD	.002*	.000*	144	6,5	9,2	12,1	44	8,2	11,5	13,1	229	4,5	7,2	10,5	886	4,0	7,7	10,9
BCE	-	-	83	-	-	-	49	-	-	-	118	-	-	-	1368	-	-	-
ANT	.000*	.706	42	13	15,3	16,2	45	7,5	11,5	14,2	22	8,1	11,9	14,7	1116	3,5	7,0	12,0

\* significatif avec  $p < .05$  après correction de Holm-Šidák (.000 signifie  $p < .001$ )

## 5.2. Analyse de lien entre questions, vote et engagement

Pour examiner la QR2 (c.-à-d. le lien entre questions/vote et engagement), nous avons comparé les populations Q vs. NQ et V vs. NV en termes d'engagement pour étudier respectivement le lien entre questions et engagement et lien entre vote et engagement en section 5.2.1. Ensuite, nous avons analysé de manière plus fine le lien entre le vote et l'engagement en section 5.2.2, notamment pour les étudiants qui posent des questions (QV vs. QNV) et ceux qui n'en posent pas (NQV vs. NQNV).

## 5.2.1. Lien entre questions et engagement et lien entre vote et engagement

### 5.2.1.1. Méthode

Pour évaluer l'engagement, nous avons considéré cette fois, pour chaque étudiant et sur chaque cours, des variables liées à l'assiduité (utilisée comme un marqueur de l'engagement, mais l'engagement est un concept plus large). Il s'agit d'une variable déclarative, établie à partir des appels faits en cours en présentiel. On en extrait 2 variables : (1) le ratio de l'assiduité globale (AssGlb) sur les deux semestres, de 0 (jamais là) à 1 (toujours là) et (2) le ratio de l'assiduité (AssCou) sur ce cours, de 0 (jamais là) à 1 (toujours là). Nous avons utilisé les mêmes tests qu'en section 5.1 (Mann-Whitney U puisque les distributions ne suivaient pas une loi normale) pour comparer les populations Q et NQ ainsi que V et NV en termes d'engagement. Nous avons effectué 2 fois 8 tests (2 comparaisons de population avec 4 cours et 2 variables à chaque fois). Nous rapportons également la taille d'effet estimée et corrigeons les seuils de significativité par la méthode de Holm-Šidák.

### 5.2.1.2. Analyse des résultats

Les résultats des tests de comparaisons de population (Q vs. NQ et V vs. NV) sur les 4 cours et avec deux variables à chaque fois (AssGlb et AssCou) sont décrits dans ce qui suit et sont résumés dans le Tableau 7. Les résultats significatifs sont mis en gras et associés à une valeur  $p < .001$  après la correction de Holm-Šidák.

En ce qui concerne les Q par rapport à NQ, pour la variable assiduité globale, il n'y avait qu'une différence statistiquement significative pour BCE ( $U = 54\,755, p < .001, \eta^2 = .024$ ). Cependant, pour l'assiduité en cours, il y avait des résultats statistiquement significatifs pour les 4 cours : BCH ( $U = 213\,974, p < .001, \eta^2 = .035$ ), HBD ( $U = 71\,554, p < .001, \eta^2 = .097$ ), BCE ( $U = 110\,005, p < .001, \eta^2 = .021$ ) et ANT ( $U = 88\,238, p < .001, \eta^2 = .042$ ).

En ce qui concerne les V par rapport à NV, nous avons trouvé également que les étudiants qui ont voté assistaient plus souvent au cours que ceux qui n'ont pas voté, sur les 4 cours considérés : BCH ( $U = 299\,389, p < .001, \eta^2 = .029$ ), HBD ( $U = 304\,339, p < .001, \eta^2 = .044$ ), BCE ( $U = 179\,322, p < .001, \eta^2 = .040$ ) et ANT ( $U = 72\,080, p < .001, \eta^2 = .036$ ) et de manière globale uniquement pour BCE ( $U = 113\,413, p < .001, \eta^2 = .008$ ).

**Tableau 7 • Caractérisation des étudiants (Q vs. NQ et V vs. NV) en termes d'engagement**

Cours	Q vs. NQ		V vs. NV	
	AssGlb	AssCou	AssGlb	AssCou
BCH	.821	.000*	.023	.000*
HBD	.113	.000*	.425	.000*
BCE	.000*	.000*	.000*	.000*
ANT	.361	.000*	.383	.000*

\* significatif avec  $p < .05$  après correction de Holm-Šidák (.000 signifie  $p < .001$ )

## 5.2.2. Relation entre vote, questions et l'engagement

### 5.2.2.1. Méthode

Pour analyser de manière plus fine le comportement des élèves en termes d'engagement, nous avons suivi la même démarche qu'en section 5.2.1 et fait des comparaisons 2 à 2 entre QV et QNV et entre NQV et NQNV, sur chacun des 4 cours, pour les deux variables AssCou et AssGlb. Pour ces 2 variables, nous avons utilisé les mêmes tests Mann-Whitney U (distributions ne suivant pas une loi normale) et effectué 2 fois 8 tests (2 comparaisons de population avec 4 cours et 2 variables à chaque fois). Les seuils de significativité ont été corrigés par la méthode de Holm-Šidák pour éviter les erreurs de type I.

### 5.2.2.2. Analyse des résultats

Les résultats des tests de comparaison des votants (QV vs. QNV et NQV vs. NQNV) en termes d'engagement (assiduité globale et assiduité en cours) et statistiques descriptives (quartiles et médiane) des 4 sous-populations sur chacun des 4 cours sont présentés dans les Tableaux 8 et 9.

En ce qui concerne les QV par rapport à QNV, aucun résultat statistiquement significatif n'a été obtenu.

En ce qui concerne les NQV par rapport à NQNV, les étudiants qui ont voté suivaient le cours plus souvent que ceux qui n'ont pas voté, sur les 4 cours considérés (comme déjà observé dans 5.2.1.2): BCH ( $U = 151652.5$ ,  $p < .001$ ,  $\eta^2 = .013$ ), HBD ( $U = 181119$ ,  $p < .001$ ,  $\eta^2 = .025$ ), BCE ( $U = 107180.5$ ,  $p < .001$ ,  $\eta^2 = .028$ ) et ANT ( $U = 25908$ ,  $p < .001$ ,  $\eta^2 = .011$ ). En revanche, aucun résultat statistiquement significatif n'a été obtenu pour l'assiduité globale.



**Tableau 8 • Caractérisation des votants en termes d'assiduité globale (AssGlb) pour les étudiants qui posent des questions (QV vs. QNV) et ceux qui n'en posent pas (NQV vs. NQNV)**

Cours	QV vs. QNV	NQV vs. NQNV	QV				QNV				NQV				NQNV			
	p	p	N	Q1	Md	Q3	N	Q1	Md	Q3	N	Q1	Md	Q3	N	Q1	Md	Q3
BCH	.106	.006	185	0,91	0,98	1	61	0,90	0,98	1	227	0,79	0,95	1	1147	0,76	0,95	1
HBD	.490	.608	154	0,93	0,98	1	54	0,93	0,98	1	262	0,88	0,96	1	1150	0,71	0,95	1
BCE	.632	.415	83	0,83	0,98	1	49	0,92	0,98	1	118	0,9	0,98	1	1368	0,74	0,95	1
ANT	.570	.691	43	1	1	1	47	0,94	1	1	23	0,98	0,98	1	1507	0,76	0,95	1

\* significatif avec  $p < .05$  après correction de Holm-Šidák (.000 signifie  $p < .001$ )

**Tableau 9 • Caractérisation des votants en termes d'assiduité en cours (AssCou) pour les étudiants qui posent des questions (QV vs. QNV) et ceux qui n'en posent pas (NQV vs. NQNV)**

Cours	QV vs. QNV	NQV vs. NQNV	QV				QNV				NQV				NQNV			
	p	p	N	Q1	Md	Q3	N	Q1	Md	Q3	N	Q1	Md	Q3	N	Q1	Md	Q3
BCH	.555	.000*	185	1	1	1	61	1	1	1	227	0,83	1	1	1147	0,33	0,83	1
HBD	.697	.000*	154	1	1	1	54	1	1	1	262	1	1	1	1150	0,5	1	1
BCE	.249	.000*	83	1	1	1	49	1	1	1	118	1	1	1	1368	0,4	1	1
ANT	.026	.000*	43	1	1	1	47	0,8	1	1	23	0,9	1	1	1507	0	0,6	1

\* significatif avec  $p < .05$  après correction de Holm-Šidák (.000 signifie  $p < .001$ )

### 5.3. Comparaison de la nature des questions posées et des questions votées

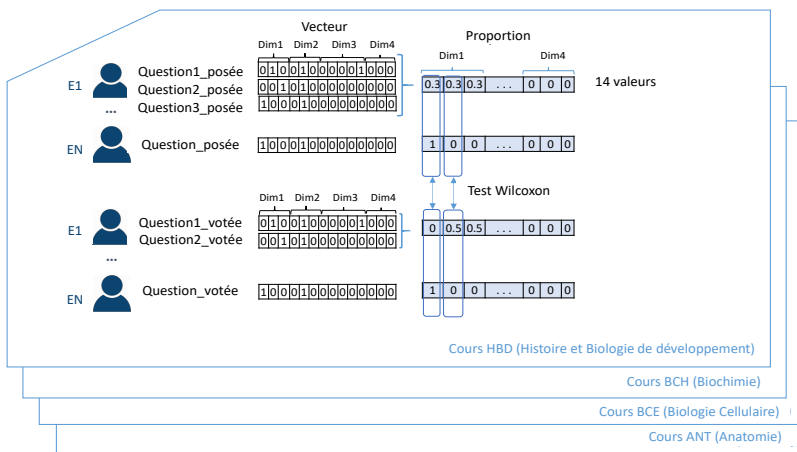
Pour examiner la QR3 (c.-à-d. pas de différences qui apparaissent entre la nature des questions posées et celle des questions votées par les étudiants), nous avons analysé la nature des questions posées et la nature des questions votées par les étudiants qui font les deux (QV).

#### 5.3.1. Méthode

Pour comparer la nature des questions que posent les étudiants à la nature des questions sur lesquelles ils votent, nous avons dû nous concentrer sur la population des étudiants qui font les deux (QV). Pour ces étudiants, sur chacun des 4 cours, nous avons considéré toutes les questions sur lesquelles ils ont voté pour calculer la proportion de chaque type de question votée dans chaque dimension. Par exemple, si dans BCH, un élève a voté sur une question de réexplication et une autre de vérification (étiquetées [Ree,0,Sch,0] et [Ver,0,0,Con]), sur la dimension 1, il aurait voté à 50 % sur des questions de valeur « Ree » (réexplication) et à 50 % sur des questions de valeur « Ver » (validation). Ces proportions sont codées entre

0 et 1, de sorte que pour chaque étudiant, sur chaque cours, on obtient un vecteur de vote composé de 14 (3+3+5+3) valeurs comprises entre 0 et 1. En suivant la même approche pour les questions posées, on peut également obtenir un vecteur de questions posées de 14 valeurs.

Une fois le prétraitement effectué, la comparaison des questions votées aux questions posées consistait à comparer pour chaque cours, pour chaque valeur d'une dimension (par ex. la valeur « Ree » de la dimension 1), la distribution de la proportion des questions posées par les étudiants et celle de questions votées dans cette dimension. En d'autres termes, comparer deux distributions (non distribuées normalement) entre 0 et 1 pour la même population d'élèves, ce qui a été fait en effectuant 14 tests de Wilcoxon, en utilisant comme en 5,1 la méthode Holm-Šidák pour ajuster la valeur  $p$  critique.



**Figure 3 • Codage des questions posées et des questions votées (par QV) en termes de proportion**

### 5.3.2. Analyse des résultats

Un seul test sur 56 a révélé un résultat statistiquement significatif (cf. Tableau 10), répondant ainsi négativement à la QR3.

**Tableau 10 • Différences entre voter et poser une question selon la nature des questions (pour QV)**

	Ree	App	Ver	Exe	Sch	Cor	Def	Man	Rai	Rol	Lie	Err	Con	Exa
BCH	.176	.427	.11	.31	.411	.017	.92	.000*	.236	.352	.295	.514	.259	.078
HBD	.382	.851	.717	.755	.809	.225	.093	.728	.007	.285	.003	.043	.941	.706
BCE	.476	.067	.015	.515	.723	.929	.89	.652	.797	.681	.118	.51	.686	.033
ANT	.826	.087	.551	.204	.795	.18	.076	.485	.212	.198	.039	.611	.691	.701

\* significatif avec  $p < .05$  après correction de Holm-Šidák (.000 signifie  $p < .001$ )

#### 5.4. Synthèse des résultats

Une synthèse de l'ensemble des résultats obtenus précédemment (cf. sections 5.1, 5.2 et 5.3) pour les différentes populations, résumant les différences significatives (jamais, parfois et toujours) pour les 4 variables considérées (note moyenne, note finale, assiduité globale et assiduité en cours), est présentée dans le Tableau 11.

**Tableau 11 • Synthèse des résultats pour les différentes populations sur les 4 variables considérées**

	Performance		Engagement	
	NotMoy	NotFin	AssGlb	AssCou
Q	-	-	+	++
NQ	-	-	-	-
V	-	+	+	++
NV	+	-	-	-
QV	+	+	-	-
QNV	-	+	-	-
NQV	-	-	-	++
NQNV	+	+	-	-

« - » : « jamais » (sur aucun cours), « + » : « parfois » (sur certains cours), « ++ » : « toujours » (sur tous les cours)

#### 6. Discussion et conclusion

Les résultats obtenus (cf. Tableau 11) révèlent plusieurs éléments intéressants qu'il convient de mettre en perspective. En termes de performance, la première analyse a révélé que le fait de poser des questions n'était pas associé à la performance des élèves (aucun résultat significatif trouvé pour la note obtenue lors de la séance de QCM en classe et la note finale). Le comportement de vote n'avait pas non plus de lien clair avec la performance : voter était négativement associé à la note moyenne pour HBD, mais positivement associé à la note finale pour ANT. Cependant, en distinguant si les étudiants votants qui posent également des questions ou non, une image plus claire apparaît : le vote est plutôt négatif, surtout lorsqu'ils n'ont pas posé de questions par ailleurs. Cette tendance semble

cependant s'inverser plus tard dans l'année, où le fait de voter en complément du fait de poser des questions entraîne de meilleurs résultats, tant aux QCM de cours qu'à l'examen final. La différence entre le premier et le second semestre pourrait être liée au fait que les étudiants les plus en difficulté ont été obligés de quitter la formation à la fin du premier semestre, ainsi qu'à la baisse générale d'activité sur la plateforme: les étudiants qui continuent à y participer sont donc probablement les plus motivés de ceux ayant suivi les cours du 1<sup>er</sup> semestre.

En ce qui concerne l'engagement, le fait de poser des questions est associé à une plus grande participation au cours (puisque les réponses sont fournies pendant le cours). La comparaison des étudiants qui votent sur des questions et ceux qui ne votent pas (V et NV) a révélé une relation positive similaire. Le vote semble être associé également aux étudiants qui sont plus susceptibles d'être présents en cours, en particulier pour les étudiants qui ne posent pas de questions et sont souvent présents au cours. Cependant, il est difficile de déterminer si les étudiants votent parce qu'ils ont l'intention d'aller au cours ou s'ils sont plus susceptibles d'y assister parce qu'ils ont voté. Il est intéressant de noter que le fait de voter et poser des questions ne semble pas lié à un engagement supérieur au fait de ne faire que l'une des deux activités. Dans notre contexte, il semble donc que des activités « actives » (au sens de Chi - ici, voter) complémentaires à des activités « constructives » (ici, poser des questions) soient plus efficaces que des activités « constructives » seules en termes d'apprentissage, mais qu'une activité « active » seule soit plus positive que de la passivité en termes d'engagement et négative en termes de performance.

L'analyse de la nature des questions posées et des questions votées par les étudiants qui font les deux (QV), c'est-à-dire les plus impliqués et ceux qui réussissent le mieux (notamment au second semestre), montre qu'il n'y a globalement pas de différence de nature entre les questions sur lesquelles ils votent et celles qu'ils posent eux-mêmes. Ce résultat est intéressant car il suggère une interprétation possible du résultat précédent, à savoir que les votes de ces étudiants correspondent effectivement bien à des questions qu'ils se posent vraiment. Il est possible que des étudiants se contentant de voter ne font pas l'effort de formuler leurs propres questions, et que s'ils le faisaient, elles seraient d'une nature différente.

Ce travail est exploratoire et présente donc plusieurs limites : même si tous les étudiants se connectent à la plateforme de questions, nous n'avons pas accès à des logs permettant de savoir s'ils ont vraiment lu les autres

questions posées. Les étudiants qui se connectent en premier n'ont également pas de questions sur lesquelles ils peuvent voter, sauf s'ils se reconnectent par la suite pour voir les nouvelles questions. Une expérience davantage contrôlée dans laquelle les étudiants doivent poser des questions et/ou voter sur des questions précédemment posées après avoir vu une vidéo, pour pouvoir passer à la suite, et complétée d'une approche qualitative (entretiens des étudiants), permettrait de vérifier les interprétations précédentes. Néanmoins le travail préalable réalisé ici rend désormais possible ce type d'expérience. Enfin, il est probable que les votes enregistrés soient en fait la manifestation de deux processus bien différents. Le premier correspond aux étudiants qui font l'effort de se poser des questions, se connectent à la plateforme et trouvant que celle-ci a déjà été posée, ne peuvent plus que la voter. Ce type de vote masque en fait une activité réelle « constructive » (au sens de Chi). Au contraire, les étudiants qui se connectent éventuellement sans question préalable, et découvrent à la lecture d'une question qu'ils se la posent également, sont dans une activité « active » (toujours au sens de Chi). Là aussi, forcer les étudiants à poser leurs questions éventuelles avant de lire celles des autres permettrait d'éviter cette ambiguïté dans le sens à donner au vote. Néanmoins il est important de souligner que le fait que des différences ont été observées entre les simples votants (les NQV qui mélangent donc des activités « actives » et « constructives ») et les poseurs de questions (les QV et QNV, entièrement dans une démarche « constructive ») plaident a priori en faveur d'un écart réel en fait encore plus important entre activités purement « actives » et purement « constructives ».

Dans notre contexte, ces résultats suggèrent qu'encourager les étudiants à formuler leurs questions, plutôt que de se contenter de voter sur les questions des autres, serait une stratégie positive pour l'apprentissage et permettrait également aux enseignants de choisir la bonne stratégie d'enseignement. En effet, il est possible que pour certains étudiants, voter donne le sentiment de faire ce qui est attendu d'eux, sans pour autant développer les stratégies métacognitives mises en jeu lorsqu'on se pose ses propres questions (identifier les concepts clés, tester sa compréhension, résumer ce qui a été appris...). Cela pourrait être fait en encourageant les étudiants à poser une question avant de pouvoir consulter celles des autres. Du point de vue des enseignants, cela signifie qu'il est d'autant plus critique de leur proposer une visualisation plus efficace que le « mur de questions » actuel pour les aider à mieux organiser leurs interventions durant la 3<sup>e</sup> semaine, ce qui a été abordé dans un autre travail (Harrak *et al.*, 2020).

Nous avons proposé des organisations alternatives de questions aux enseignants *via* un questionnaire pour évaluer l'utilisabilité de nos propositions et particulièrement le schéma de codage développé. Envisager des tableaux de bord personnalisés pour les enseignants à partir des organisations proposées est l'une des perspectives principales de ce travail.

En résumé, voter est une bonne stratégie pour les étudiants sachant déjà formuler leurs propres questions. En revanche, pour ceux en difficulté, cela peut retarder la prise de conscience de leurs lacunes et leur capacité à les combler activement.

Globalement, notre schéma de codage permettrait d'annoter les questions des étudiants de manière plus fine en termes d'intentions et de nourrir la réflexion de l'enseignant pour lui proposer éventuellement une réaction pédagogique différente sur les questions posées. L'annotation automatique de questions permettrait également d'identifier des caractéristiques du profil des étudiants en termes de performance et d'autres aspects de leur comportement. Il est donc important de noter que notre processus d'annotation et le schéma de codage utilisé pour les questions posées par les étudiants de PACES, dans le cadre d'une classe inversée, devraient pouvoir être facilement répliqués et réutilisés dans d'autres contextes et travaux.

## **REMERCIEMENTS**

Nous remercions Pierre Gillois de nous avoir fourni les données.

## **RÉFÉRENCES**

Abdi, H. (2007). Bonferroni and Sidak corrections for multiple comparisons. *Encyclopedia of Measurement and Statistics*, 3, 103-107.

Artstein, R. et Poesio, M. (2008). Inter-coder agreement for computational linguistics. *Computational Linguistics*, 34(4), 555-596. <https://doi.org/10.1162/coli.07-034-R2>

Bergsma, W. (2013). A bias-correction for Cramér's V and Tschuprow's T. *J Korean Statistical Society*, 42(3), 323-328. <https://doi.org/10.1016/j.jkss.2012.10.002>

Bihani, A., Ullman, J. D. et Paepcke, A. (2018). FAQtor : Automatic FAQ generation using online forums. Dans K.E. Boyer et M. Yudelson (dir.), *Proceedings of the 11<sup>th</sup> International Conference on Educational Data Mining (EDM 2018)* (p. 529-532). ERIC.

Bloom, B. S. et Engelhart, M. B., Furst, E. J., Hill, W. H. et Krathwohl, D. R. (1956). *Taxonomy of educational objectives. The classification of educational goals. Cognitive domain: 1*. Addison-Wesley Longman Ltd.

(Chi et Wylie, 2014)

Chi, M. T. H. et Wylie, R. (2014). The ICAP framework: linking cognitive engagement to active learning outcomes. *Educational Psychologist*, 49(4), 219-243. <https://doi.org/10.1080/00461520.2014.965823>

Chin, C. et Brown, D. E. (2002). Student-generated questions: a meaningful aspect of learning in science. *International Journal of Science Education*, 24(5), 521-549. <https://doi.org/10.1080/09500690110095249>

Chin, C. et Kayalvizhi, G. (2002). Posing problems for open investigations: what questions do pupils ask? *Research in Science & Technological Education*, 20(2), 269-287. <https://doi.org/10.1080/0263514022000030499>

Fritz, C. O., Morris, P. E. et Richler, J. J. (2012). Effect size estimates: current use, calculations, and interpretation. *Journal of Experimental Psychology: General*, 141(1), 2-18. <https://doi.org/10.1037/a0024338>

Graesser, A. C. et Person, N. K. (1994). Question asking during tutoring. *American educational research journal*, 31(1), 104-137.

Harper, K. A., Etkina, E. et Lin, Y. (2003). Encouraging and analyzing student questions in a large physics course: meaningful patterns for instructors. *Journal of Research in Science Teaching*, 40(8), 776-791. <https://doi.org/10.1002/tea.10111>

Harrak, F., Bouchet, F., Luengo, V. et Gillois, P. (2020). Evaluating teachers' perceptions of students' questions organization. Dans C. Rensing et H. Drachler (dir.), *Proceedings of the 10<sup>th</sup> International Conference on Learning Analytics & Knowledge (LAK 2020)* (p. 11-16). ACM. <https://doi.org/10.1145/3375462.3375509>

Harrak, F., Bouchet, F. et Luengo, V. (2019). From students' questions to students' profiles in a blended learning environment. *Journal of Learning Analytics*, 6(1), 54-84. <https://doi.org/10.18608/jla.2019.61.4>

Harrak, F., Bouchet, F., Luengo, V. et Gillois, P. (2018). Profiling students from their questions in a blended learning environment. Dans A. Pardo K. Bartimote-Aufflick (dir.), *Proceedings of the 8<sup>th</sup> International Conference on Learning Analytics and Knowledge (LAK 2018)* (p. 102-110). ACM. <https://doi.org/10.1145/3170358.3170389>

Jiang, Z., Zhang, Y., Liu, C. et Li, X. (2015). Influence analysis by heterogeneous network in MOOC forums: what can we discover? Dans O. C. Santos, J. G. Boticario, C. Romero, M. Pechenizkiy, A. Merceron, P. Mitros (dir.), *Proceedings of the 8<sup>th</sup> International Conference on Educational Data Mining (EDM 2015)* (p. 242-249). ERIC.

Kiss, T. et Strunk, J. (2006). Unsupervised multilingual sentence boundary detection. *Comput Linguist*, 32(4), 485-525.

Li, H., Duan, Y., Clewley, D. N., Morgan, B., Graesser, A. C., Shaffer, D. W. et Saucerman, J. (2014). Question asking during collaborative problem solving in an online game environment. Dans S. Trausan-Matu, K. E. Elizabeth Boyer, M. Crosby, K. Panourgia (dir.), *Proceedings of the 12<sup>th</sup> International Conference on Intelligent Tutoring Systems (ITS 2014)* (p. 617-618). Springer.

MacFarland, T. W. et Yates, J. M. (2016). Mann-Whitney U Test. Dans T. W. MacFarland et J. M. Yates (dir.), *Introduction to Nonparametric statistics for the biological sciences using R* (p. 103-132). Springer International Publishing. [https://doi.org/10.1007/978-3-319-30634-6\\_4](https://doi.org/10.1007/978-3-319-30634-6_4)

Otero, J. et Graesser, A. C. (2001). PREG: Elements of a model of question asking. *Cognition and instruction*, 19(2), 143-175.

Sagot, B. et Fišer, D. (2008). Building a free French wordnet from multilingual resources. *OntoLex*. <https://hal.inria.fr/inria-00614708/document>

Scardamalia, M. et Bereiter, C. (1992). Text-based and knowledge based questioning by children. *Cognition and Instruction*, 9(3), 177-199. [https://doi.org/10.1207/s1532690xci0903\\_1](https://doi.org/10.1207/s1532690xci0903_1)

Sullins, J., McNamara, D., Acuff, S., Neely, D., Hildebrand, E., Stewart, G. et Hu, X. (2015). Are you asking the right questions: The use of animated agents to teach learners to become better question askers. Dans I. Russell and W. Eberle (dir.). *Proceedings of the 28<sup>th</sup> International Florida Artificial Intelligence Research Society Conference (FLAIRS 2015)* (p. 479-481). AAAI Press. <https://asu.pure.elsevier.com/en/publications/are-you-asking-the-right-questions-the-use-of-animated-agents-to->

Supraja, S., Hartman, K., Tatinati, S. et Khong, A. W. (2017). Toward the automatic labeling of course questions for ensuring their alignment with learning outcomes. Dans X. Hu, T. Barnes, A. Hershkovitz et L. Paquette (dir.), *Proceedings of the 10<sup>th</sup> International Conference on Educational Data Mining (EDM 2017)* (p. 56-63). ERIC.

Teixeira-Dias, J.J. C., Pedrosa de Jesus, H., Neri de Souza, F. et Watts, M. (2005). Teaching for quality learning in chemistry. *International Journal of Science Education*, 27(9), 1123-1137. <https://doi.org/10.1080/09500690500102813>

Wong, J.-S., Pursel, B., Divinsky, A. et Jansen, B.J. (2015). An analysis of MOOC discussion forum interactions from the most active users. Dans N. Agarwal, K. Xu et N. Osgood (dir.), *Proceedings of the 8<sup>th</sup> International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction (SBP 2015)* (p. 452-457). Springer.

Zeng, Z., Chaturvedi, S. et Bhat, S. (2017). Learner affect through the looking glass: Characterization and detection of confusion in online courses. Dans X. Hu, T. Barnes, A. Hershkovitz et L. Paquette (dir.), *Proceedings of the International Conference on Educational Data Mining (EDM 2017)* (p. 272-277). ERIC.





# Une approche hybride à la modélisation de l'apprenant dans un STI pour l'apprentissage du raisonnement logique

► **Roger NKAMBOU, Ange TATO** (CRIA, Université du Québec à Montréal), **Janie BRISSON, Serge ROBERT** (LANCI, Université du Québec à Montréal), **Maxime SAINTE-MARIE** (Aarhus, Université du Danemark)

---

---

■ **RÉSUMÉ** • Cet article présente une démarche hybride pour la modélisation de l'apprenant dans le STI MUSE-Logique conçu pour l'apprentissage du raisonnement logique. Un réseau bayésien à validation experte est d'abord proposé, suivi par un modèle construit à partir des données qui permet d'améliorer la structure du réseau bayésien et son initialisation pour un nouvel apprenant. Un autre modèle construit par apprentissage profond sur des données, complète la solution hybride proposée. La combinaison de ces trois méthodes confère au modèle apprenant une validité induite des connaissances expertes et des données, et améliore son pouvoir prédictif.

■ **MOTS-CLÉS** • Modèle apprenant, diagnostic/traçage de connaissances.

■ **ABSTRACT** • *This paper presents a hybrid approach to learner modeling into MUSE-Logic, an ITS for logical reasoning learning. A Bayesian network built by experts is first proposed, followed by a model built from data that improves both the Bayesian network structure and its initialization for a new learner. A second model built from deep learning on data, completes the proposed hybrid solution. The combination of these three methods gives the learner model both expert and data validity, and significantly improves its predictive power.*

■ **KEYWORDS** • *Student model, cognitive diagnosis, knowledge tracing.*

## **1. Introduction**

De nombreuses expériences en sciences cognitives montrent que des erreurs systématiques sont courantes chez l'humain dans l'usage du raisonnement logique (Giroto *et al.*, 1989 ; Rossi et Van der Henst, 2007). Un certain nombre de questions se posent lors de la recherche de solutions pour améliorer les compétences humaines dans ce domaine : quels sont les phénomènes impliqués dans l'acquisition de compétences de raisonnement logique ? Peut-on éliciter ces compétences ainsi que les connaissances sous-jacentes ? Quelles sont les stratégies d'adaptation pour favoriser le développement de la compétence logique ? Quelles sont les caractéristiques d'un système de tutorat intelligent (STI) pour soutenir cet apprentissage ? Les réponses à ces questions ne peuvent être apportées sans une explication et une compréhension appropriée des connaissances sous-tendant le raisonnement logique et les erreurs commises par les humains, couplées avec la participation active d'experts, notamment des logiciens, des psychologues du raisonnement et des professionnels de l'éducation en logique.

Le projet Muse-Logique regroupe des chercheurs dans une équipe multidisciplinaire mise en place en 2013. L'objectif est d'étudier les bases de l'acquisition des compétences de raisonnement logique, de comprendre les difficultés liées à cet apprentissage et de créer un STI capable de détecter, de diagnostiquer et de corriger les erreurs de raisonnement dans diverses situations. Pour ce faire, notre approche de conception participative vise à :

- fournir un catalogue des erreurs de raisonnement logique ;
- développer une théorie originale des causes de ces erreurs, y compris une théorie de la compétence en raisonnement logique ;
- analyser les structures formelles du traitement de l'information que l'on trouve dans les systèmes logiques ;
- développer un système de tutorat intelligent pour aider les apprenants à améliorer leurs capacités de raisonnement ;
- ultimement, aider les experts à redéfinir la nature de la rationalité humaine en utilisant les STI comme un banc d'essai pour tester d'autres hypothèses de la science cognitive du raisonnement.

L'approche participative adoptée a mené à l'élaboration de composants essentiels pour la construction du STI Muse-Logique. Ceux-ci ont été préalablement élicites, validés et argumentés par les experts. Ces composants incluent le catalogue des erreurs, les structures et métastructures du raisonnement logique et les stratégies remédatives

favorisant l'émergence de bons modèles mentaux. Muse-Logique dans sa version actuelle implémente ces composants pour la logique classique des propositions et offre une panoplie d'activités permettant à un apprenant de développer sa compétence du raisonnement logique dans plusieurs classes de situations que nous avons établies avec les experts : concrètes, contrefactuelles et abstraites.

Cet article décrit la démarche de conception participative de Muse-Logique, notamment les fonctionnalités des trois modules issus de celle-ci. La section 2 décrit les fondements théoriques sous-jacents à Muse-Logique, suivie par les sections 3 et 4 qui décrivent les résultats de la conception participative des trois principaux composants du système : le modèle du domaine, le modèle de l'apprenant et le tuteur. Une attention particulière est accordée à l'approche hybride de la modélisation de l'apprenant, approche qui allie réseau bayésien, modèle de diagnostic cognitif et apprentissage profond pour offrir un modèle apprenant doté d'un meilleur pouvoir prédictif. La section 5 illustre l'implémentation de Muse-Logique, elle est suivie d'une conclusion qui résume les contributions et quelques travaux futurs.

## **2. Cadre théorique**

Pourquoi construire un système pour le raisonnement logique ? Cette idée part de différents constats listés ci-dessous.

- Une partie importante de la connaissance humaine relève du traitement de l'information par des raisonnements (Markovits, 2014).
- Le raisonnement logique occupe une place importante dans la cognition (Evans, 2002).
- Le raisonnement logique humain est plus souvent fondé sur les logiques non classiques (comme les logiques intuitionnistes) que classiques (Robert, 2005).
- En tant que machine cognitive en situation de lutte pour sa survie, l'humain tend à faire des erreurs systématiques dans ses raisonnements logiques (Evans, 2002 ; Evans *et al.*, 1993).
- Ces erreurs peuvent être dues à l'emploi de stratégies de découverte (inductives, analogiques, abductives) dans un contexte de justification.
- Apprendre à raisonner logiquement, c'est apprendre comment fonctionne le traitement cognitif de l'information, tout en apprenant les lois et procédures valides du raisonnement logique (Robert, 2005).

**Roger NKAMBOU, Ange TATO, Janie BRISSON, Serge ROBERT,  
Maxime SAINTE-MARIE**

- Les mécanismes cognitifs du raisonnement humain (Mercier *et al.*, 2017) sont le produit de modules cérébraux relativement autonomes, mais également en interaction étroite entre eux.

À partir de ces postulats dérivant de la littérature en psychologie du raisonnement, l'objectif de nos travaux a consisté à rassembler les développements récents dans les domaines des systèmes tutoriels intelligents, des sciences cognitives et de l'intelligence artificielle pour proposer un environnement d'apprentissage capable d'aider les étudiants à améliorer leurs compétences en logique. Notre perspective pour étudier le raisonnement logique est :

- naturaliste, les mécanismes du raisonnement sont le résultat d'une histoire évolutionnaire qui peut être expliquée (Dennett, 2003) ;
- cognitiviste, la cognition est du traitement d'information ;
- méliorativiste, la compréhension des normes logiques peut nous aider à améliorer nos compétences en raisonnement logique.

Dans ce contexte, l'information est définie comme de la régularité, par opposition à la probabilité. Le raisonnement consiste à tirer de l'information (conclusion), à partir d'informations données (prémisses), en appliquant une règle (implicite ou explicite). Le type de règle définit la nature du raisonnement. Quand ces règles de raisonnement sont logiques, elles sont non ampliatives, c'est-à-dire qu'elles n'augmentent pas l'information entre les prémisses et la conclusion, par opposition aux raisonnements créatifs, qui sont ampliatifs. Ainsi, notre hypothèse sur l'erreur logique est que lorsqu'elle est syntaxique, elle vient du traitement du raisonnement logique comme s'il s'agissait d'un raisonnement créatif. C'est le cas des sophismes (un des deux types d'erreurs de raisonnement que nous traitons dans ce travail). Les sophismes sont en effet des inférences qui créent de l'information sans justification pour le faire (Robert, 2005). Les erreurs surviennent aussi lorsque les contenus alternatifs ne sont pas considérés. Le deuxième type d'erreur que nous traitons dans nos travaux est connu comme étant « la suppression d'inférence » (Espino et Byrne, 2020). Nous exploitons des théories liées à ces deux types d'erreurs pour offrir un cadre d'apprentissage du raisonnement logique capable de tracer ces erreurs et d'aider l'apprenant à les inhiber. Par exemple, des études ont prouvé que la sensibilité aux contre-exemples était un excellent moyen d'inhiber les sophismes de l'implication (Markovits *et al.*, 2013).

Les STI ont fait avancer la compréhension de l'apprentissage en général, mais ont aussi fait leurs preuves quant à leur capacité à soutenir

L'apprentissage du raisonnement en sciences, mathématiques, et même en logique (Barnes et Stamper, 2010 ; Lesta et Yacef, 2002 ; Tchetagni *et al.*, 2007). Nous pouvons formuler deux limites des STI proposés en logique. D'abord, ces systèmes ne se fondent pas sur une élicitation des connaissances et des structures logiques essentielles au raisonnement, limitant ainsi leur portée explicative du raisonnement de l'apprenant. Ensuite, ils ne se fondent pas sur des théories fondamentales de l'erreur de raisonnement au centre desquelles se trouve la théorie des processus duaux (Evans *et al.*, 1993 ; Stanovich, 2011). Celle-ci identifie deux types de processus de raisonnement qui peuvent être activés dans le raisonnement logique : des processus de type 1, spontanés, automatiques, inconscients, peu coûteux sur le plan des ressources cognitives et probablement innés, et des processus de type 2, réfléchis, récents, contrôlés, et qui font appel aux ressources cognitives telle la mémoire de travail. Les processus de type 1 nous permettent de faire certains raisonnements logiquement valides, mais ils sont aussi responsables de plusieurs erreurs. Ces processus de type 1 ne peuvent pas être éliminés, ils sont trop ancrés en nous, mais nous pouvons apprendre à les inhiber et à activer des processus de type 2. Cette activation accroît considérablement notre compétence logique en nous permettant d'éviter les erreurs causées par les processus de type 1. Nous appuyant sur ces thèses, nous explorons les différences entre ces deux types de processus et les stratégies pour inhiber les premiers et activer les seconds. Cette perspective s'arrime avec la théorie de l'inhibition cognitive de Houdé, qui indique l'importance d'apprendre à inhiber les automatismes acquis (qui s'exercent grâce au système 1, si on fait un parallèle avec la théorie des processus duaux). Ces automatismes acquis constituent la source de plusieurs erreurs de raisonnement tant chez l'enfant que chez l'adulte (Houdé, 2000 ; Houdé, 2019).

L'activation des processus de type 2 et la correction de l'erreur logique ne peuvent se faire sans l'inhibition des processus de type 1 et cette inhibition exige de prendre conscience de l'erreur. En voulant améliorer la théorie du développement cognitif de Piaget, Moshman (Moshman, 2004) a démontré qu'une prise de distance métacognitive est essentielle à cette prise de conscience de l'erreur et à son inhibition. Notre projet explore cette thèse et tente de montrer comment des détours métacognitifs sont essentiels à l'acquisition de la compétence logique. Ainsi, lorsque plongée dans différents contextes, mais avec un même problème soumis, une personne éprouverait de la difficulté à bien conclure, même dans une logique apparemment simple comme la logique des propositions, elle

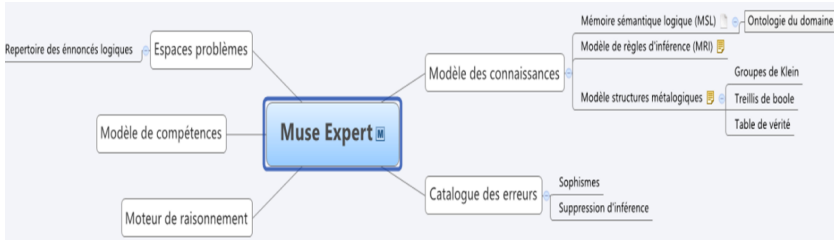
pourrait réussir dans un contexte plus commun qu'un contexte plus abstrait. Le contexte ici renvoie à la manière dont est formulé le problème. Il peut être familier (causal), contrefactuel ou encore abstrait. Les théories évolutionnistes ont pu faire ressortir l'influence des biais et des heuristiques liés aux effets de contexte. Ces théories soutiennent que même si nous avons la technique logique appropriée pour raisonner, certains facteurs tels que les biais de croyance et les heuristiques avec lesquels nous évoluons toute notre vie influenceront notre raisonnement (Cosmides, 1989 ; Varin, 2007). Grâce à ces théories, il devient possible d'expliquer certains raisonnements invalides.

Selon Robert, le processus d'opérations mentales met en relation le raisonnement et la catégorisation de l'information (Robert, 2005). Pour raisonner, il faut organiser et structurer l'information en catégories, établir des liens entre elles, sélectionner et appliquer des règles logiques nécessaires sur ces informations. La conception de notre modèle du domaine Muse-Expert a fait appel à la mise en œuvre de ces théories. La catégorisation, dans ce cas, réfère à l'organisation de la connaissance, la distinction des règles d'inférences (valides et invalides), la décomposition des informations de chaque problème sur lesquels le système doit raisonner et la représentation des structures/modèles dont ce dernier est doté.

### **3. Conception participative du module expert**

Sur le plan architectural, Muse-Logique est constitué des trois modules classiques d'un système tutoriel intelligent (Nkambou *et al.*, 2010) : un module expert, un module tuteur et un modèle de l'apprenant. L'environnement d'apprentissage donne accès à une banque de problèmes de raisonnement logique générés dans des situations variées. Chaque module offre une interface qui permet sa manipulation individuelle par un utilisateur.

Comme présenté dans la figure 1, le module Muse-Logique Expert met en œuvre aussi bien les habiletés et connaissances en raisonnement logique que les mécanismes de raisonnement connexes (syntaxique et sémantique des règles du système logique donné).



**Figure 1 • Composants du module Muse-Logique Expert**

En fait, puisque nous prévoyons de couvrir plusieurs systèmes logiques, Muse-logique Expert comporte un contrôleur général dont le but est de sélectionner la logique la plus pertinente dans laquelle la situation (de l'énoncé du problème actuel) s'adapte. Pour ce faire le système s'appuie sur des méta-connaissances pour rechercher des motifs (marqueurs de chaque système logique fournis par les experts) dans la situation qui offre une meilleure prédiction, conduisant à la sélection du système logique approprié. Dans cet article, nous nous concentrons uniquement sur un système logique (la logique classique des propositions).

### **3.1. Spécification de la mémoire sémantique et procédurale de la logique propositionnelle**

La conception participative de chaque composant du module expert a été soigneusement réalisée dans l'équipe. Tout d'abord, nous avons étudié le domaine de la logique propositionnelle ce qui nous a amenés à la spécification complète de tous les concepts qui lui sont liés, puis conduits à une ontologie de domaine que nous avons implémentée comme un modèle OWL. L'ontologie formelle a ensuite été validée lors d'une séance de travail avec les experts logiciens.

L'équipe a ensuite entrepris d'éliciter le processus de déclenchement des connaissances procédurales. Cette démarche a abouti à la spécification de toutes les règles d'inférence logique dans la logique propositionnelle y compris les règles valides et non valides. Huit règles d'inférence valides ont été spécifiées, deux pour chacun des quatre opérateurs binaires de base de la logique propositionnelle (conjonction, disjonction, implication et incompatibilité). Durant ce processus, les experts logiciens et psychologues du raisonnement ont travaillé ensemble sur la validation empirique de certaines preuves liées à des concepts logiques « mal définis » tels que les constructions d'incompatibilité. L'étude des huit autres règles non valides spécifiées a conduit à l'élaboration du catalogue des erreurs de

raisonnement. Les erreurs de raisonnement ont été divisées en deux catégories: les sophismes (inférences invalides) et les suppressions d'inférences valides.

Les sophismes sont des erreurs de raisonnement logique dans lesquelles le raisonneur ne parvient pas à trouver le bon lien entre les prémisses et la conclusion des règles en cause. Quant aux suppressions d'inférences valides, elles se produisent quand il ou elle ajoute de nouvelles informations dans les prémisses rendant impossible la conclusion.

### **3.2. Élaboration des métastructures du raisonnement en logique propositionnelle**

L'équipe a examiné en détail l'ancrage sémantique du raisonnement à travers les structures métalogiques comme les treillis de Boole ou encore les groupes de Klein. En fait, le raisonnement logique a été considéré à la fois dans ses composantes syntaxique et sémantique. Quand la logique associe ainsi le syntaxique et le sémantique, la notion de structure devient centrale. Un système logique est fait de structures. Une structure est un ensemble d'objets doté de relations (comme  $>$ ,  $=$  ...) et/ou d'opérations (comme  $+$ ,  $x...$ ). Une structure qui permet de composer de nouveaux objets par des opérations est une structure algébrique, tandis qu'une structure de relations entre des objets est une structure d'ordre. Chaque structure formelle se distingue par ses connecteurs ou ses opérateurs et par les propriétés que possèdent ces relateurs ou opérateurs. La logique classique est régie par des structures logiques telles que l'algèbre de Boole et une structure d'ordre appelée « treillis de Boole ». Les logiques non classiques peuvent être obtenues par des affaiblissements ou des élargissements de l'algèbre de Boole ou du treillis de Boole. Dans cette perspective, l'apprentissage de différents systèmes logiques peut être vu comme étant l'acquisition de différentes structures algébriques pour catégoriser le monde et de différentes structures d'ordres pour établir des hypothèses de dépendance causale entre ces catégories. Nous traitons alors les stratégies de type 1 causant les erreurs logiques comme des simplifications des structures à l'œuvre dans les raisonnements. D'un point de vue métacognitif, nous utilisons ces métastructures pour illustrer le raisonnement erroné, interprété comme un mauvais parcours de celles-ci ou comme un écrasement de structure (groupe de Klein) (Robert et Brisson, 2016).



### 3.3. Des situations de raisonnement multiples

Le raisonnement logique n'est pas un processus absolu dans le sens que le même type de raisonnement peut être difficile à appliquer d'une classe de situations à une autre. Être compétent dans l'usage du modus ponendo ponens (MPP) dans une situation donnée (par exemple en situation descriptive causale avec peu de possibilités) n'implique pas qu'on le sera dans une autre (par exemple en situation descriptive causale avec beaucoup d'alternatives). Cette thèse est soutenue par une étude menée par les membres de l'équipe Muse-Logique dans laquelle certains sophismes étaient attribuables à la difficulté de trouver des contre-exemples dans un raisonnement mettant en œuvre l'affirmation du conséquent (Brisson *et al.*, 2014 ; Markovits, 2014). Nous avons examiné, avec les experts logiciens et les psychologues du raisonnement membres de l'équipe, différentes familles de situation pouvant engendrer une variation du niveau de difficulté d'un type de raisonnement. Nous avons ainsi établi six familles de situations descriptives (Tableau 1).

**Tableau 1 • Familles de situations de raisonnement (classes de problèmes)**

DESCRIPTIVE				NORMATIVE
Causal		Abstract		
Concret	Contrary-to-fact			
Few Alternatives (DCCFA)	Few Alternatives (DCCFFA)	Formal (AF)	Informal (AI)	
Many Alternatives (DCCMA)	Many Alternatives (DCCFMA)			

La catégorisation fournit non seulement une base permettant de classer les compétences en raisonnement, mais permet aussi d'organiser les activités d'apprentissage du raisonnement. Le raisonnement normatif a été exclu parce que les expérimentations du raisonnement sur du contenu normatif ont montré une excellente performance tant des adultes que des enfants (Cloutier, 2016 ; Girotto *et al.*, 1989). Une explication de cet effet de contenu est que la cognition humaine comporte un module spécialisé pour raisonner sur les normes sociales (Girotto *et al.*, 1989).

La catégorisation nous a permis d'établir une mémoire procédurale de la logique des propositions comportant 96 éléments de connaissances, soit 48 éléments de connaissances valides et 48 non valides. Par ailleurs cette contextualisation aide à classer les exercices à soumettre à l'apprenant et définit un cadre précis pour l'évaluation des compétences de l'apprenant. Par

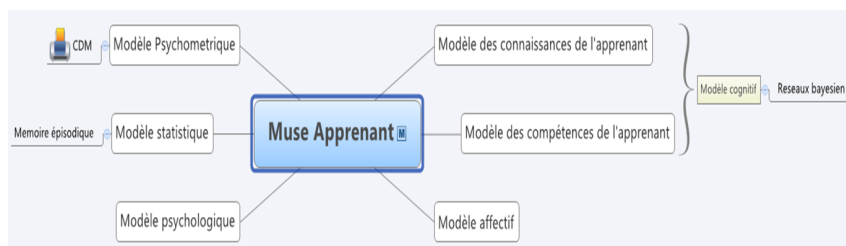
exemple, la catégorie « Descriptive causale et concrète avec peu d’alternatives » (DCCFA) comporte des exercices avec un contenu familier, facile à comprendre, et dont l’antécédent présente très peu d’alternatives. Par exemple, pour l’antécédent « si un chien a des puces, alors il se gratte constamment », il pourrait y avoir deux alternatives : « avoir la peau sèche » ou « une maladie de la peau ». Par contre, dans la catégorie « Causale avec beaucoup d’alternatives », en plus d’un contenu familier, facile à comprendre, il y a beaucoup d’alternatives causales. Par exemple, l’antécédent de « Si on lance une pierre dans la fenêtre, alors la fenêtre se brisera » a un grand nombre d’alternatives : « lancer une chaise dans la fenêtre », « lancer une brique dans la fenêtre », « collision avec une voiture », « tempête tropicale », « frapper la fenêtre avec un bâton », « se projeter dans la fenêtre », etc.

Dans Muse-Logique, les exercices sont présentés sous forme de syllogisme et offrent des choix de réponse. Par exemple :

- Règle : Si je lance une pierre dans la fenêtre, alors la fenêtre brisera.
- Information : La fenêtre est brisée.
- Que peut-on conclure ?
  - On a lancé une pierre dans la fenêtre.
  - On n’a pas lancé une pierre dans la fenêtre.
  - On ne peut conclure.

#### **4. Le modèle de l’apprenant**

Le modèle de l’apprenant dans Muse-Logique comporte plusieurs dimensions (figure 2). La mémoire épisodique garde la trace des exercices effectués par l’apprenant ainsi que tous les éléments de performance reliés. Le modèle cognitif est principalement un modèle bayésien dont les nœuds sont les 96 unités de connaissances liées au raisonnement. C’est de l’approche de construction du modèle cognitif dont nous parlerons dans la suite de cet article.



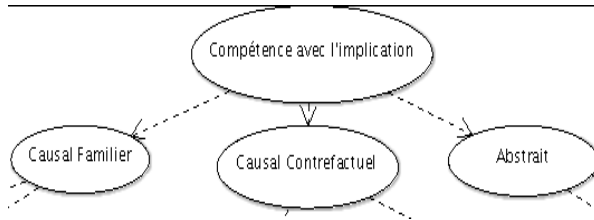
**Figure 2. • Composants du modèle de l’apprenant dans Muse-Logique**

Lorsqu'un réseau bayésien (RB) est utilisé pour la modélisation d'un apprenant, des nœuds observables représentent un ensemble d'items et des nœuds cachés (ou latents) représentent une compétence du domaine d'apprentissage. Les liens entre les nœuds représentent la dépendance bayésienne entre la réussite d'un item et une compétence du domaine, entre deux compétences ou entre deux items. Le modèle de l'apprenant généré par cette structure contient donc deux parties : le modèle de la compétence et le modèle de la performance (Almond *et al.*, 2007). Ce formalisme est tout à fait adéquat pour représenter les compétences pertinentes au raisonnement déductif ainsi que les performances aux différents items. Pour le modèle de l'apprenant de Muse-Logique, nous avons donc construit la structure d'un RB en tant qu'outil d'évaluation automatique du raisonnement conditionnel. Les probabilités a priori ainsi que les relations d'influence entre les différentes unités de connaissance ont été initialement établies par les experts. Certains nœuds sont directement connectés aux exercices de raisonnement (items). Certains nœuds du réseau sont bien entendu liés aux inférences invalides (sophismes ou suppressions d'inférence). On entretient ainsi et de manière automatique le modèle de l'apprenant sur chaque unité de connaissances (valide ou erronée). Les statistiques sur la performance sont calculées à partir des données stockées dans la mémoire épisodique.

#### **4.1. Un réseau bayésien initial à validation experte**

La structure à validation experte que nous avons construite et l'identification des compétences visées pour une maîtrise complète du raisonnement conditionnel sont basées sur les effets du contenu dans le raisonnement, sur l'approche développementale de Markovits (Markovits, 2014), ainsi que sur l'approche formelle de Robert et Brisson (Robert et Brisson, 2016). Les compétences visées sont l'inhibition des conditions contraignantes, la génération des antécédents alternatifs ainsi que la bonne gestion mentale des trois situations pertinentes pour l'implication (i.e. les trois rangées vraies du connecteur). Ces trois compétences de base sont déclinées dans trois situations différentes, soit le raisonnement avec du contenu familier, contrefactuel ou abstrait. Les nœuds au bas de la structure représentent les types d'items qui seront sélectionnés par le système, soit les quatre formes logiques du raisonnement conditionnel.

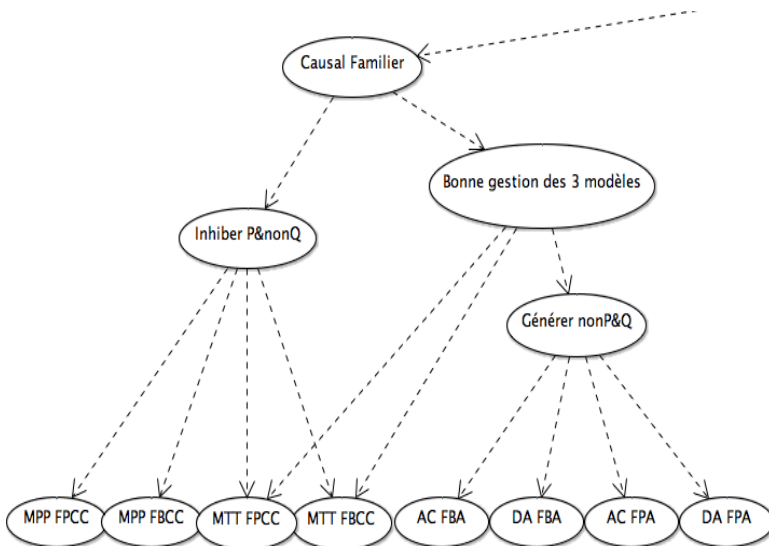
Nous exposons ci-dessous les différentes portions de la structure du RB pour le raisonnement conditionnel. La figure 3 montre le haut de la structure. Le premier nœud représente la compétence globale. La structure montre des liens de dépendance entre ce dernier et les trois nœuds qui lui succèdent.



**Figure 3 • Tête du RB pour le raisonnement conditionnel**

En lisant de gauche à droite, la première branche du RB expose les compétences pertinentes pour le raisonnement conditionnel avec du contenu de plus en plus difficile. Le premier nœud enfant du nœud racine « Compétence avec l'implication » est donc « Causal Familier », suivi de « Causal contrefactuel » et de « Abstrait ».

La figure 4 expose les compétences pertinentes pour le raisonnement conditionnel avec du contenu familial.



**Figure 4 • Portion du RB pour le raisonnement conditionnel familial**

Dans cette structure, le nœud « Inhiber P&nonQ » représente la capacité d'inhiber les conditions contraignantes de la prémisse conditionnelle, nécessaire pour reconnaître la certitude des inférences valides MPP et MTT. Les nœuds enfants de ce dernier sont donc les nœuds qui représentent le MPP avec peu (MPP FPCC) et beaucoup (MPP FBCC) de conditions contraignantes ainsi que le MTT avec peu (MTT FBCC) et beaucoup (MTT FPCC) de conditions contraignantes.

De plus, le nœud « Bonne gestion des 3 modèles » représente une compétence pertinente aux raisonnements MTT, AC et DA. Cette compétence consiste à garder en mémoire de travail les trois modèles pertinents à une prémisse conditionnelle ou, autrement dit, les trois rangées vraies du connecteur d'implication. Le MPP, de par sa simplicité et de par le fait qu'il peut être réussi même avec une compréhension simplifiée de l'implication, n'est pas un bon indicateur de cette compétence avancée. Nous avons donc déterminé qu'une capacité à répondre correctement aux trois inférences plus complexes (MTT, AC et DA) était le meilleur indicateur de cette compétence.

Le nœud « Générer nonP&Q » représente la capacité de générer les antécédents alternatifs qui serviront de contre-exemples aux conclusions invitées des inférences AC et DA. On retrouve donc parmi ses nœuds enfants ceux qui représentent la compétence avec le AC et le DA avec beaucoup (AC FBA et DA FBA respectivement) et peu (AC FPA et DA FBA respectivement) d'antécédents alternatifs. Le Tableau 2 résume la signification de chacun de ces nœuds et donne un exemple pour chacun d'eux.

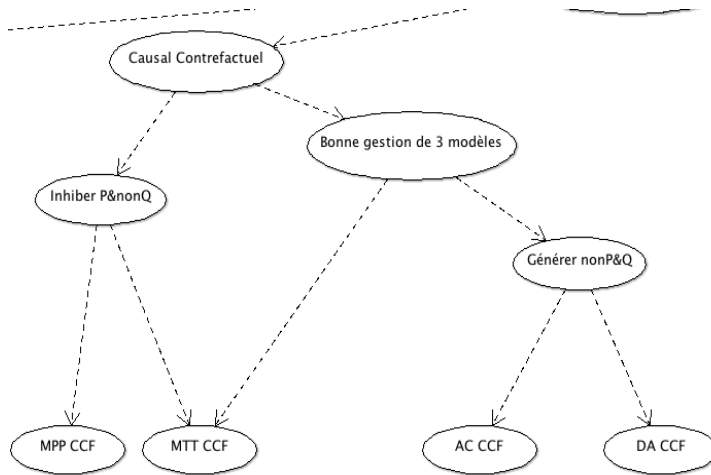
**Tableau 2 • Légende pour les nœuds items familiaux**

<b>Nœud</b>	<b>Signification</b>	<b>Exemple</b>
<b>MPP</b>	<i>Modus Ponendo Ponens</i> Familier	Si Sébastien saute dans la piscine, alors il sera mouillé.
<b>FPCC</b>	Peu de Conditions Contraignantes	Sébastien saute dans la piscine. Donc, Sébastien est mouillé.
<b>MPP</b>	<i>Modus Ponendo Ponens</i> Familier	Si la clé de contact a été tournée, alors la voiture démar-
<b>FBCC</b>	Beaucoup de Conditions Contraignantes	ra. La clé de contact a été tournée. Donc, la voiture démarrera.
<b>MTT</b>	<i>Modus Tollendo Tollens</i> Familier	Si Sébastien saute dans la piscine, alors il sera mouillé.
<b>FPCC</b>	Peu de Conditions Contraignantes	Sébastien n'est pas mouillé. Donc, Sébastien n'a pas sauté dans la piscine
<b>MTT</b>	<i>Modus Tollendo Tollens</i> Familier	Si la clé de contact a été tournée, alors la voiture démar-
<b>FBCC</b>	Beaucoup de Conditions Contraignantes	ra. La voiture n'est pas démarrée. Donc, la clé de contact n'a pas été tournée.
<b>AC</b>	Affirmation du Conséquent Familier	Si on lance une pierre dans une fenêtre, alors la fenêtre se
<b>FBA</b>	Beaucoup d'Alternatives	brisera. La fenêtre est brisée. Donc, on a lancé une pierre dans la fenêtre.
<b>DA</b>	Déni de l'Antécédent Familier	Si on lance une pierre dans une fenêtre, alors la fenêtre se
<b>FBA</b>	Beaucoup d'Alternatives	brisera. On n'a pas lancé de pierre dans la fenêtre. Donc, la fenêtre n'est pas brisée.
<b>AC</b>	Affirmation du Conséquent Familier	Si un chien a des puces, alors il se gratte constamment.
<b>FPA</b>	Peu d'Alternatives	Fido se gratte constamment. Donc, Fido a des puces.
<b>DA</b>	Déni de l'Antécédent Familier	Si un chien a des puces, alors il se gratte constamment.
<b>FPA</b>	Peu d'Alternatives	Fido n'a pas de puces. Donc, Fido ne se gratte pas constamment.

Finalement, les relations de dépendance entre les nœuds de compétences exposés permettront au système de déterminer les probabilités de maîtrise des compétences à partir du patron de réponses de l'apprenant. Plus précisément, un ensemble de bonnes réponses aux items MPP et MTT augmentera la probabilité de maîtriser l'inhibition des conditions contraignantes. Des bonnes réponses aux items AC et DA augmenteront la probabilité de maîtriser la génération des antécédents alternatifs. De plus, un ensemble de bonnes réponses aux MTT, AC et DA augmentera la probabilité d'une bonne gestion des trois modèles pertinents à l'implication. L'évaluation de toutes ces compétences aura un impact sur la probabilité de maîtrise du raisonnement conditionnel avec du contenu familial, ce qui, finalement, aura un impact sur la probabilité de maîtriser le raisonnement conditionnel dans son ensemble.

La figure 5 expose la portion du RB relative au raisonnement avec du contenu contrefactuel. Les nœuds ainsi que les relations de dépendance entre eux sont les mêmes que dans la structure relative au raisonnement

avec du contenu familier : la structure réfère aux mêmes processus de raisonnement et aux mêmes compétences visées. La différence est au niveau des nœuds qui représentent les formes logiques. Notons que la distinction entre peu et beaucoup d’alternatives ou de conditions contraignantes n’est plus présente dans cette structure. En effet, la quantité d’antécédents alternatifs et de conditions contraignantes avec des prémisses contrefactuelles est difficilement mesurable et aucune étude n’investigue cette dimension dans la littérature actuelle. Nous nous sommes donc restreints à des nœuds qui représentent la compétence au raisonnement avec les quatre formes logiques avec du contenu contrefactuel, sans distinction supplémentaire.



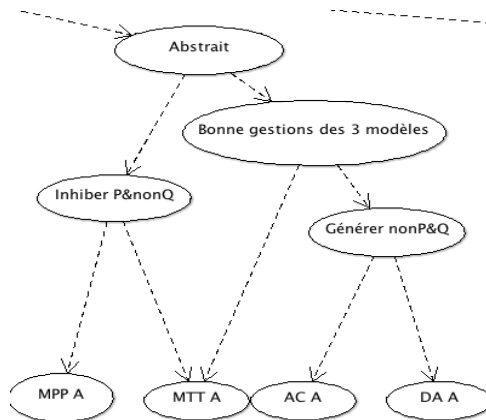
**Figure 5 • Portion du RB pour le raisonnement causal contrefactuel**

Le Tableau 3 résume la signification de chacun de ces nœuds avec un exemple pour chacun.

**Tableau 3 • Légende pour les nœuds items causaux  
contrefactuels**

Nœud	Signification	Exemple
<b>MPP</b> <b>CCF</b>	MPP Contrefactuel	Si on lance du ketchup sur une chemise, alors elle deviendra propre. On a lancé du ketchup sur une chemise. Donc, la chemise est propre.
<b>MTT</b> <b>CCF</b>	MTT Contrefactuel	Si Louise se brosse les dents régulièrement, alors elle aura des caries. Louise n'a pas de caries. Donc, Louise ne se brosse pas les dents régulièrement.
<b>AC</b> <b>CCF</b>	AC Contrefactuel	Si on lance du ketchup sur une chemise, alors elle deviendra propre. Une chemise est propre. Donc, on a lancé du ketchup sur la chemise.
<b>DA</b> <b>CCF</b>	DA Contrefactuel	Si Louise se brosse les dents régulièrement, alors elle aura des caries. Louise ne se brosse pas les dents régulièrement. Donc, Louise n'a pas de caries.

La figure 6 expose la dernière portion du RB, soit le raisonnement avec du contenu abstrait. Les nœuds « Inhiber P&nonQ », « Bonne gestion des 3 modèles » et « Générer nonP&Q » sont présents, entretiennent les mêmes relations de dépendance et réfèrent à des processus de raisonnement semblables à ceux des structures précédentes. Les nœuds « MPP A », « MTT A », « AC A » et « DA A » réfèrent aux quatre formes logiques avec du contenu abstrait.



**Figure 6 • Portion du RB pour le raisonnement causal abstrait**



Notons que, comme dans le cas du raisonnement contrefactuel, la distinction quant au nombre d'antécédents alternatifs et de conditions contraignantes n'est pas présente. La raison est toutefois différente. À ce niveau de difficulté du raisonnement conditionnel, le raisonneur doit comprendre que, pour toute prémisse conditionnelle, un antécédent alternatif peut être généré, et ce, indépendamment de son bagage de connaissances. La génération des alternatives, l'inhibition des conditions contraignantes et la gestion des trois modèles demandent donc, à ce niveau, une compréhension plus abstraite que dans les situations précédentes.

Le Tableau 4 fournit des exemples pour des nœuds qui représentent les quatre formes logiques avec du contenu abstrait (abrégié A).

**Tableau 4 • Légende pour les nœuds items causaux abstraits**

Nœud	Exemple
<b>MPP A</b>	Si une personne morp, alors elle deviendra plede. Pierre morp. Donc, Pierre deviendra plede.
<b>MTT A</b>	Si une personne morp, alors elle deviendra plede. Pierre n'est pas devenu plede. Donc, Pierre ne morp pas.
<b>ACA A</b>	Si une personne morp, alors elle deviendra plede. Pierre est devenu plede Donc, Pierre morp.
<b>DA A</b>	Si une personne morp, alors elle deviendra plede. Pierre ne morp pas. Pierre n'est pas devenu plede.

Finalement, la structure globale du RB rend compte de la hiérarchie entre les compétences appartenant aux trois niveaux de contenu. Ainsi, les compétences relatives au raisonnement abstrait demandent une certaine maîtrise des compétences relatives au raisonnement contrefactuel, qui elles-mêmes demandent une certaine maîtrise des compétences relatives au raisonnement familier.

#### **4.2. Amélioration de la structure et du processus d'initialisation du réseau bayésien par le CDM**

Allié au réseau bayésien construit avec les experts, un modèle de diagnostic cognitif de type Cognitive Diagnosis Model (CDM) (Rupp *et al.*, 2007) a été entraîné à partir des données collectées lors d'une expérimentation avec 294 participants sur une banque de 48 exercices.

**Roger NKAMBOU, Ange TATO, Janie BRISSON, Serge ROBERT, Maxime SAINTE-MARIE**

Pour ce faire, une Q-Matrice (dont une portion est présentée dans le Tableau 5) mettant en évidence le croisement entre les items (48 items regroupés en 16 classes) et les unités de connaissances ou habiletés cognitives (9 classes) a été soigneusement élaborée avec l'aide des experts logiciens.

**Tableau 5 • Q-Matrice dans MUSE-Logique**

	Causal Familiar			Causal Contrary to Fact			Abstract		
	Inhibit	Generate	Manage	Inhibit	Generate	Manage	...	...	...
MPP FMD	1	0	0	0	0	0	0	0	0
MTT FFD	1	0	1	0	0	0	0	0	0
DA CCF	0	1	1	0	1	1	0	0	0
...	...	...	...	...	...	...	...	...	...
MTT A	1	0	1	1	0	1	1	0	1
AC A	0	1	1	0	1	1	0	1	1
DA A	0	1	1	0	1	1	0	1	1

Nous avons ensuite entraîné un modèle DINA (plus précisément G-DINA) (de la Torre, 2009 ; de la Torre, 2011) sur les 294 patrons de réponses, ce qui a permis d'estimer des paramètres utiles comme *guess*, la probabilité qu'un apprenant puisse répondre correctement à un exercice sans avoir les habiletés nécessaires, et *slip*, pour la probabilité d'une mauvaise solution alors qu'on disposait des habiletés nécessaires, ou encore l'indice de discrimination associé à chaque item de la Q-Matrice qui permet de déterminer leur pertinence dans la mesure de la compétence ciblée. Les sous-sections suivantes présentent quelques aperçus de l'analyse des résultats.

**4.2.1. Qualité de l'ajustement du modèle**

Bien que l'évaluation de l'ajustement de modèle se fasse généralement via le critère d'information Akaike (AIC) et le critère d'information bayésien (BIC), un indicateur absolu de la qualité de l'ajustement pour le modèle CDM est la mesure  $\chi^2$ , test du « khi-deux » par paire d'items (Chen *et al.*, 2013 ; Chen *et al.*, 2018). Cette mesure indiquera que le modèle est inadéquat si la valeur-*p* de la mesure  $\chi^2$  maximale est supérieure à 0,01, le seuil de signification asymptotique fixée selon la littérature (Groß *et al.*, 2016). Dans notre cas, le  $\chi^2$  maximal = 33,95 avec  $p = 6,793\ 31610^{-7}$  (largement < 0,01), confirmant ainsi que le modèle obtenu est adéquat.

### 4.2.2. Probabilités marginales des compétences

Le Tableau 6 ci-dessous montre la probabilité moyenne de maîtrise pour chaque compétence.

**Tableau 6 • Distribution des compétences**

Estimate	Skill Distribution									
	InhibitFam	GenerateFam	ManageFam	InhibitCF	GenerateCF	ManageCF	InhibitAbs	GenerateAbs	ManageAbs	$\alpha$
	0,9772	0,9606	0,9706	0,4427	0,5512	0,5393	0,6762	0,6365	0,6428	0,71

La compétence la plus difficile semble être « Inhibit counterfactual », avec une probabilité moyenne de 44,3 % de maîtrise. En effet, cette compétence exige d'inhiber les conditions invalidantes aux énoncés conditionnels contrefactuel, ce qui n'est pas facile.

Cette observation corrobore le fait que l'indice de discrimination d'un item (IDI) le plus élevé est celui du MPP contrefactuel comme on peut le voir dans le Tableau 7. Rappelons que l'IDI mesure la capacité d'un item à faire la distinction entre les répondants compétents et ceux qui ne le sont pas (selon la compétence sous-jacente). Plus cet indice est faible, plus l'item est pertinent pour la mesure de cette compétence.

Par ailleurs, la probabilité beaucoup plus élevée pour les compétences avec le niveau familial suggère une séparation avec les deux autres niveaux, avec une différence entre le niveau contrefactuel et le niveau abstrait.

### 4.2.3. L'estimation des paramètres *guess*, *slip* & IDI

Tout d'abord, comme on peut le voir dans le Tableau 7, nous avons noté une estimation élevée du *guess* et une faible IDI pour les deux MTT avec un contenu familial. Cela pourrait s'expliquer par le fait qu'une interprétation biconditionnelle de la prémisse majeure conduit à la bonne réponse au MTT alors même que la compétence avec le conditionnel n'est pas acquise. Cependant, dans l'ensemble, tous les autres éléments ont une bonne IDI. Il convient également de noter que l'IDI du MPP contrefactuel est la plus élevée.

Le succès à cet item semble donc le meilleur moyen de montrer que le raisonnement conditionnel est maîtrisé. Cela va dans le sens des observations tirées des probabilités marginales des compétences discutées plus haut

**Tableau 7 • Probabilités Guess, Slip et IDI**

	Guess		Slip		Item Easin.	Item Discrim.
	Est.	SE	Est.	SE	(G+(1-S ))/2	1-G-S
MPPFewDis	0,27	0,04	0,02	0,01	0,06	0,72
MPPManyDis	0,19	0,03	0,05	0,02	0,05	0,76
MPPCCF	0,00	0,00	0,05	0,02	0,04	0,95
MPPAbs	0,17	0,02	0,00	0,00	0,02	0,83
MTTFewDis	0,51	0,06	0,09	0,02	0,15	0,40
MTTManyDis	0,45	0,05	0,14	0,03	0,14	0,41
MTTCF	0,07	0,01	0,19	0,03	0,14	0,74
MTTAbs	0,22	0,03	0,09	0,02	0,15	0,69
ACMAnyAlt	0,08	0,02	0,03	0,01	0,03	0,89
ACFewAlt	0,02	0,01	0,25	0,05	0,10	0,73
ACCCF	0,02	0,01	0,07	0,02	0,05	0,91
ACAbs	0,01	0,00	0,11	0,03	0,07	0,87
DAMAnyAlt	0,10	0,02	0,04	0,02	0,05	0,86
DAFewAlt	0,03	0,01	0,20	0,05	0,03	0,77
DACCF	0,11	0,02	0,04	0,02	0,15	0,85
DAAbs	0,05	0,02	0,04	0,02	0,12	0,91
<b>α</b>	0,14	0,02	0,09	0,02	0,08	0,77

#### 4.2.4. Corrélations tétrachoriques

Un des résultats de l'approche CDM est l'estimation des corrélations entre les habiletés. Le Tableau 8 illustre les corrélations établies par notre modèle.

**Tableau 8 • Corrélations entre compétences**

	Skill Correlations									α
	InhibitFam	GenerateFam	ManageFam	InhibitCF	GenerateCF	ManageCF	InhibitAbs	GenerateAbs	ManageAbs	
InhibitFam	1	0,85	0,9	-0,06	0,05	0,04	0,19	0,15	0,15	0,36
GenerateFam	0,85	1	0,78	0,05	0,06	0,16	0,2	0,16	0,17	0,38
ManageFam	0,9	0,78	1	0,22	0,06	0,04	0,2	0,15	0,16	0,39
InhibitCF	-0,06	0,05	0,22	1	0,64	0,6	0,67	0,5	0,52	0,46
GenerateCF	0,05	0,06	0,06	0,64	1	0,15	0,31	0,4	0,35	0,34
ManageCF	0,04	0,16	0,04	0,6	0,15	1	0,35	0,41	0,43	0,35
InhibitAbs	0,19	0,2	0,2	0,67	0,31	0,35	1	0,3	0,31	0,39
GenerateAbs	0,15	0,16	0,15	0,5	0,4	0,41	0,3	1	0,34	0,38
ManageAbs	0,15	0,17	0,16	0,52	0,35	0,43	0,31	0,34	1	0,38
<b>α</b>	0,36	0,38	0,39	0,46	0,34	0,35	0,39	0,38	0,38	0,38

Selon la taille de notre échantillon, les corrélations supérieures à 0,33 sont considérées comme significatives pour un seuil de coefficient de corrélation  $\alpha = 0,05$  (Guilford et Lyons, 1942). Nous avons donc appliqué ce critère dans notre analyse subséquente en illustrant aussi le coefficient de corrélation moyen ( $\alpha$ ) de chaque compétence.

Une observation que l'on peut tirer de ces corrélations tétrachoriques est que les compétences ayant un contenu familial sont fortement

corrélées à d'autres compétences du même niveau de contenu, alors que les corrélations avec les niveaux contrefactuel et abstraits ne sont pas significatives. Cependant, les compétences au niveau contrefactuel sont corrélées entre elles et avec les compétences au niveau abstrait, et vice-versa. Dans l'ensemble, cela semble indiquer une fois de plus une séparation entre le niveau familier du contenu et les deux autres niveaux.

#### **4.2.5. Autres résultats et impacts sur le réseau bayésien initial**

Plusieurs autres résultats importants du modèle ne sont pas rapportés dans cet article. C'est par exemple le cas de la matrice *posterior* qui associe les patrons de réponses aux profils de compétence des apprenants, offrant ainsi la possibilité d'estimer le vecteur de compétence initial d'un nouvel apprenant, utilisé pour l'initialisation du réseau bayésien représentant son modèle cognitif. Ainsi, les probabilités a priori du réseau bayésien initial sont estimées de cette manière plutôt que de considérer la chance. Par ailleurs, le modèle CDM que nous avons construit pour MUSE-Logique a permis de valider et d'ajuster la structure du réseau initialement construit avec les experts, aboutissant ainsi à une structure plus efficace informée à la fois par les données (*bottom-up*) et les experts (*top-down*).

#### **4.3. Évaluation et prédiction des performances par le BKT et le DKT**

La modélisation de l'apprenant dans un système visant l'apprentissage passe avant tout par la modélisation de sa connaissance, puisque c'est cette dernière que le système vise à améliorer. L'approche la plus populaire pour modéliser la connaissance de l'utilisateur est le *Knowledge Tracing* (KT) (pour traçage des connaissances). Le KT vise à modéliser la façon dont les connaissances des apprenants évoluent pendant l'apprentissage. La connaissance est modélisée sous forme d'une variable latente et est mise à jour en fonction des performances de l'utilisateur-apprenant au fur et à mesure qu'il effectue les tâches. Cette approche est formalisée comme suit :

Sachant les interactions d'un apprenant jusqu'au temps  $t$  ( $x_1... x_t$ ) sur une tâche d'apprentissage particulière, quelle performance va-t-il accomplir au temps  $t+1$ , l'objectif étant d'estimer la probabilité  $p(x_{t+1}|x_1..., x_t)$  ? Il existe plusieurs solutions pour estimer cette probabilité entre autres le *Bayesian Knowledge Tracing* (BKT) et le *Deep Knowledge Tracing* (DKT).

### **4.3.1. Pallier les limites du BKT et du DKT classiques par une approche hybride**

Nous avons montré dans les sections précédentes comment le réseau bayésien de Muse-Logique a été construit et validé. Nous expliquons dans cette section comment le BKT y est appliqué. Rappelons que le BKT (Corbett et Anderson, 1994) est une approche de modélisation latente de l'apprenant par un réseau bayésien qui peut être soit appris à partir des données soit défini par les experts. Le BKT est un cas particulier des chaînes de Markov cachées où les connaissances de l'utilisateur sont représentées par un ensemble de variables binaires (la connaissance est maîtrisée ou non). Les observations sont aussi binaires : un usager peut soit réussir soit échouer à résoudre un problème (Yudelson *et al.*, 2013). Il ne peut pas réussir à 30 % ou à 70 % par exemple. Cependant, il existe une certaine probabilité ( $G$ , paramètre *guess*) que l'utilisateur donne une réponse correcte à un exercice  $X$  sachant qu'il ne maîtrise pas la connaissance liée à cet exercice. Pareillement, un usager qui maîtrise une connaissance donnera généralement une réponse correcte sur les exercices liés à cette connaissance, mais il existe une certaine probabilité ( $S$ , le paramètre *slip*) que l'utilisateur n'y réponde pas correctement. Le modèle standard de BKT est donc défini par quatre paramètres : l'état de connaissance initiale, la vitesse d'apprentissage (*learning parameters*), les probabilités de *slip* et de *guess* (*mediating parameters*). En règle générale, le BKT utilise les informations a priori sur les connaissances, telles que les probabilités conditionnelles et les probabilités de maîtrise des connaissances, pour mesurer les progrès de l'apprentissage des usagers. Dans un tel contexte, la performance (échec ou réussite) est la variable observée et les connaissances sont les variables latentes. L'inférence bayésienne est utilisée pour déterminer les probabilités recherchées. Le BKT est classique et, au-delà de Muse-Logique, a été utilisé avec succès dans plusieurs domaines d'apprentissage, entre autres la programmation informatique (Kasurinen et Nikula, 2009) et la lecture (Beck et Chang, 2007).

Toutefois, l'utilisation d'un réseau bayésien implique parfois de définir manuellement les probabilités a priori et d'étiqueter manuellement les interactions avec des concepts appropriés. De plus, les données de réponses binaires utilisées pour modéliser les connaissances, les observations et les transitions imposent une limite aux types d'exercices pouvant être modélisés. Le DKT a donc été proposé afin de résoudre les problèmes du BKT. Sans nécessairement entrer dans une description détaillée du DKT

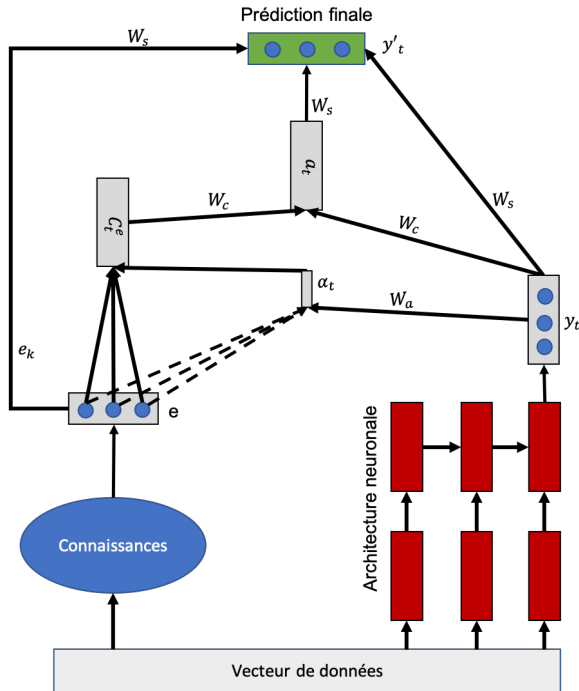
original (Piech *et al.*, 2015), notons que deux limites lui sont facilement reconnues, notamment :

- sa faible capacité à bien généraliser sur les connaissances plancher/plafond (c'est-à-dire celles avec peu de données d'entraînement, par exemple un faible taux d'échec ou un taux élevé de succès) ;
- la non intégration des connaissances expertes, une source importante déjà éprouvée dans le cadre des BKT à validation experte.

Ces deux limites nous ont poussés à développer un DKT amélioré dit hybride (Tato et Nkambou, 2019) qui a été utilisé dans le cadre de cette étude. Une des améliorations a porté sur l'injection des connaissances expertes à travers un mécanisme attentionnel tel qu'illustré dans la figure 7.

En d'autres termes, le réseau de neurones « consultera » ou « portera attention sur » les connaissances expertes avant de prendre la décision finale (émettre la prédiction finale ( $y't$ )). Les « connaissances » peuvent être représentées sous plusieurs formes. Dans le cas de MUSE-Logique (cf. section 4.3.2), elles sont incarnées par le réseau bayésien conçu intégralement par les experts. L'importance que le modèle neuronal accordera à ce que ces connaissances prédisent comme sortie est calculée au moyen de pondération de l'attention ( $W_a$  et  $W_c$ ).

Au fur et à mesure de l'apprentissage, le réseau saura quelle importance il accordera à chacune des prédictions provenant de la branche des connaissances, en fonction des essais/erreurs qu'il aura commis. Les poids  $W_a$  représentent l'importance de chaque caractéristique apprise par l'architecture neuronale ( $y_i$ ) par rapport à chaque caractéristique extraite à partir des connaissances. Les poids  $W_c$  représentent l'importance des prédictions faites à partir des connaissances (via le vecteur de contexte) et des caractéristiques apprises ( $y_i$ ) pour l'estimation du vecteur de prédiction finale. Ainsi, le modèle se concentrera sur ce que dit l'expertise dont les connaissances a priori et a posteriori (intégrées dans le réseau bayésien qui modélise l'apprenant), avant de prendre une décision. En fusionnant ainsi les connaissances des experts avec l'architecture neuronale à l'aide du mécanisme d'attention, le modèle traite de façon itérative les connaissances en sélectionnant le contenu pertinent à chaque étape.



**Figure 7 • Modèle DKT hybride**

À chaque instant  $t$ , le modèle déduit, par une couche de concaténation, l'état caché attentionnel (le vecteur  $\alpha_t$ ). Cette couche combine la prédiction actuelle du système neuronal ( $y_t$ ) et le vecteur de contexte côté expert ( $C_t^e$ ) pour déduire le vecteur attentionnel  $\alpha_t$  comme suit :

$$\alpha_t = \tan(W_c[C_t^e; y_t])$$

Le vecteur de contexte  $C_t^e$  est calculé comme étant la moyenne pondérée, selon  $\alpha_t$  sur chaque entrée du vecteur de connaissances expert ( $e$ )<sup>1</sup> comme suit :

---

<sup>1</sup> Le vecteur  $e$  rend compte de l'état des connaissances de l'apprenant tel que prédit par les experts, obtenu par inférence bayésienne sur le réseau bayésien initial (construit par les experts du domaine).



$$C_t^e = \sum_k \alpha_{t,k} \cdot e$$

avec

$$\alpha_{t,k} = \frac{\exp(\text{score}(e_k, y_t))}{\sum_{j=1}^s \exp(\text{score}(e_j, y_t))}$$

et

$$\text{score}(e_k, y_t) = e_k y_t W_a + b$$

Dans ces formules, la variable  $e$  représente un vecteur de longueur égale à la taille du vecteur à prédire. Chaque élément  $e_k$  du vecteur  $e$  représente la probabilité de maîtrise de la  $k$ -ième connaissance selon les prédictions de l'expert (c'est à dire du réseau bayésien).

$1 \leq k \leq s$ ,  $e$  est la prédiction faite à partir des connaissances a priori et a posteriori,  $y_t$  est la prédiction actuelle faite par l'architecture neuronale et  $s$  est la taille du vecteur prédit (le nombre de classes à prédire).

Le score est un vecteur basé sur le contenu qui calcule la corrélation (score d'alignement) entre les connaissances expertes et les caractéristiques latentes apprises par l'architecture neuronale, tout en considérant le biais ( $b$ ).

Ainsi, le score définit comment les connaissances expertes et les caractéristiques apprises à partir des données sont alignées. Le modèle attribue un score  $\alpha_{t,k}$  à la paire d'entités à la position  $t$  et aux connaissances expertes  $(e_k, y_t)$ , en fonction de leur correspondance. L'ensemble des  $\alpha_{t,k}$  sont des pondérations définissant à quel point chaque caractéristique de la donnée provenant de l'expert doit être prise en compte pour chaque sortie (prédiction finale).

Le système (architecture) neuronal est un *DKTm*, une version du DKT original dans laquelle nous utilisons un masque pour obliger une attention particulière de l'apprentissage du réseau sur les connaissances plancher/plafond (pour une meilleure généralisation sur ces connaissances). Ce principe est détaillé dans (Tato et Nkambou, 2019). En bref, l'injection de la connaissance experte dans le réseau de neurones se fait par un mécanisme attentionnel induit par le vecteur  $\alpha_t$  résultant d'une combinaison des vecteurs de sortie des sous-systèmes neuronal et bayésien ( $y_t$  et  $C^e$ ), tel qu'illustré dans la figure 7.

### **4.3.2. Prédiction des performances par le DKT dans MUSE-Logique**

Le traçage des connaissances dans MUSE-Logique par le BKT appliqué sur la version du réseau bayésien à validation experte uniquement (sans aucune amélioration par le CDM ou toute autre approche utilisant les données) produit des déductions avec une précision relative intéressante (*Accuracy* de 65 %). Notre hypothèse est qu'en utilisant ce réseau bayésien comme connaissance experte (a priori) nous pouvons mettre en place un nouveau modèle de traçage de connaissances hybride plus performant, qui le fusionne avec le DKT. Pour ce faire, nous avons conduit des tests qui visent à évaluer la capacité de traçage de connaissance du DKT hybride versus le réseau bayésien seulement.

Nous avons utilisé les mêmes données collectées dans le cadre de notre étude sur l'amélioration du réseau bayésien par le CDM (section 4.2). Dans notre ensemble de données, chaque ligne de données représente un participant (un total de 294 données et une longueur de séquence de 48). Cette quantité de données est trop limitée pour entraîner un modèle d'apprentissage profond. Cependant, en les combinant aux connaissances des experts, nous voyons une différence substantielle dans les résultats. Les exercices ont été encodés en utilisant les 16 connaissances (compétences) de base identifiées, ce qui signifie que les questions relatives à la même connaissance sont encodées avec le même Id (1~16). Les connaissances avec peu de données d'entraînement (connaissances planchers/plafonds) sont déterminées par une comparaison de la moyenne des bonnes réponses obtenues pour chaque connaissance. Dans le Tableau 9, nous avons fait la moyenne de toutes les réponses pour chaque connaissance. Les connaissances planchers sont celles dont les valeurs moyennes sont les plus basses et les connaissances plafonds sont celles dont les moyennes sont les plus élevées.

La deuxième et la dernière partie de notre nouvelle fonction de perte portent donc sur ces connaissances. Puisque le RNN n'accepte qu'une longueur fixe de vecteurs en entrée, nous avons utilisé un codage *one-hot* pour convertir la performance des apprenants en des vecteurs de longueur fixe dont tous les éléments sont à 0 sauf un seul qui est à 1. Le 1 dans ce vecteur indique deux choses : quelle connaissance a été évaluée et si la réponse à la question portant sur cette connaissance est correcte.

**Tableau 9 • Distribution des réponses par rapport aux connaissances difficiles à maîtriser (moyenne<0,4) et connaissances faciles à maîtriser (moyenne>0,9)**

Connaissances	N	Moyenne	Écart-type
MppFd	294	<b>0,94</b>	0,16
MppMd	294	0,89	0,23
MppCcf	294	<b>0,90</b>	0,23
MppA	294	<b>0,96</b>	0,16
MttFd	294	0,84	0,26
MttMd	294	0,79	0,29
MttCcf	294	0,74	0,33
MttA	294	0,84	0,28
AcMa	294	0,42	0,38
AcFa	294	<b>0,30</b>	0,36
AcCcf	294	<b>0,33</b>	0,40
AcA	294	<b>0,28</b>	0,41
DaMa	294	0,40	0,37
DaFa	294	<b>0,30</b>	0,35
DaCcf	294	<b>0,38</b>	0,40
DaA	294	<b>0,30</b>	0,42

Nous avons utilisé 3 modèles : le DKT, le DKT où nous avons appliqué un masque à la fonction de perte (DKTm) (Tato et Nkambou, 2019), et le DKTm hybride avec connaissances a priori (DKTm+BN) décrit à la figure 7. Nous avons utilisé AAdam, une version accélérée de Adam (Tato et Nkambou, 2020), comme algorithme d'optimisation de l'apprentissage. Nous avons utilisé 20 % des données pour les tests et 15 % pour la validation.

Comme mentionné ci-dessus, le BN à lui seul a une performance de 65 % de précision globale sur ces données. Le résultat est évalué à l'aide de la métrique F1score sur chaque connaissance (traitée comme 2 classes : réponses correctes et incorrectes) en jeu et sur la précision globale du modèle. Les modèles ont été évalués sur 20 expériences différentes et la moyenne des résultats finaux a été calculée. Le paramètre de régularisation a été fixé à 0,10 dans toutes ces expériences. Notre implémentation du modèle DKT avec Keras a été inspirée de l'implémentation faite par Khajah, Lindsey et Mozer (Khajah *et al.*, 2016).

### **4.3.3. Analyse des résultats du DKT dans MUSE-Logique**

Nous avons tout d'abord observé les résultats en suivant les interactions pour 1 apprenant. Ensuite, nous avons étudié les interactions avec l'ensemble des étudiants et avons itéré 20 fois afin de considérer la moyenne des prédictions des différents modèles. L'analyse des performances de la prédiction pour 1 étudiant est présentée à la section 4.3.3.1 alors que l'analyse des performances prédictives moyennes des modèles pour l'ensemble des apprenants est présentée à la section 4.3.3.2.

#### ***4.3.3.1. Prédiction pour un seul apprenant***

Les figures A1, A2 et A3 en annexe illustrent une visualisation du traçage en temps réel des connaissances en raisonnement logique d'un apprenant. Dans la visualisation, nous ne montrons que les prédictions dans le temps pour les 48 exercices. Les colonnes représentent la compétence mise en jeu dans la question posée à l'instant  $t$ , suivie de la réponse réelle de l'apprenant (il y a 48 lignes représentant les 48 questions et 16 colonnes pour les 16 compétences de base en jeu). Les couleurs et les probabilités à l'intérieur de chaque case indiquent la probabilité prédite par les modèles que l'apprenant répondra correctement à une question liée à une compétence (en ligne) à l'instant suivant. Plus la couleur est foncée, plus la probabilité est élevée.

L'apprenant répond d'abord correctement à une question relative à la connaissance DA\_FFA (DA\_FFA -1) et à une question relative à la compétence AC\_FMA (AC\_FMA -1) et ensuite échoue à la question portant sur la connaissance MTT\_FDD (MTT\_FDD -0). Dans les 45 questions suivantes, l'apprenant résout une série de problèmes liés aux 16 connaissances impliquées.

Dans la figure A1, on constate facilement que le DKT est incapable de faire des prédictions précises sur les compétences avec peu de données, surtout celles qui sont difficiles à maîtriser.

Les figures A2 et A3 illustrent la prédiction des performances du même apprenant avec les 2 variantes de DKT que nous avons proposées. Dans ce scénario, l'apprenant a donné 2 bonnes réponses sur 3 sur la compétence AC\_FMA (5<sup>e</sup> colonne) et nous voyons que le DKT<sub>m</sub> et le DKT<sub>m</sub>+BN sont capables de mieux prédire cette information, comparativement au DKT.

Notre première version du DKT (DKTm, figure A2) surpasse le DKT initial (figure A1) pour le traçage de toutes les connaissances en jeu. De plus, le DKTm amélioré avec la connaissance experte (le réseau bayésien) en utilisant notre solution hybride (DKTm+BN, figure A3) surpasse tous les autres modèles par sa capacité de prédiction avec peu de données.

**4.3.3.2. Analyse du pouvoir prédictif des modèles pour l'ensemble des apprenants**

Chaque fois qu'un apprenant répond à un exercice, les modèles mettent à jour ses connaissances de façon à pouvoir prédire s'il répondra correctement ou non à un exercice portant sur chaque connaissance lors de sa prochaine interaction. Les résultats de la performance des trois variantes de DKT après 20 itérations sont présentés dans les Tableaux 10 et 11

**Tableau 10 • Le DKT, le DKTm et le DKKm+BN sur les connaissances difficiles**

F1score	AC_FMA		DA_FMA		AC_FFA		DA_FFA		AC_CCF		DA_CCF		AC_A		DA_A		Acc
DKT	0.74	0.09	0.72	0.0	0.79	0.0	0.79	0.0	0.79	0.0	0.73	0.0	0.80	0.0	0.79	0.0	0.74
DKTm	0.70	0.36	0.36	0.55	0.82	0.37	0.80	0.47	0.83	0.54	0.76	0.51	0.85	0.48	0.86	0.46	0.80
<b>DKTm+BN</b>	<b>0.70</b>	<b>0.64</b>	<b>0.64</b>	<b>0.43</b>	<b>0.86</b>	<b>0.74</b>	<b>0.80</b>	<b>0.47</b>	<b>0.87</b>	<b>0.80</b>	<b>0.82</b>	<b>0.72</b>	<b>0.83</b>	<b>0.58</b>	<b>0.88</b>	<b>0.80</b>	<b>0.82</b>

Dans le Tableau 10, la dernière colonne est la précision globale du modèle. Pour chaque connaissance, la première colonne correspond à la valeur du f1score pour la prédiction des réponses incorrectes et la deuxième colonne pour la prédiction des réponses correctes. La meilleure valeur pour chaque connaissance est en gras.

**Tableau 11 • Le DKT, le DKTm et le DKKm+BN sur les éléments de connaissances faciles**

F1score	MPP_FFD		MPP_FMD		MTT_FFD		MTT_FMD		MPP_CCF		MTT_CCF		MPP_A		MTT_A	
DKT	0.0	0.97	0.0	0.94	0.0	0.91	0.0	0.90	0.0	0.96	0.0	0.87	0.0	0.96	0.0	0.90
DKTm	0.0	0.97	0.02	0.97	0.0	0.94	0.22	0.89	0.0	0.95	0.09	0.84	0.0	0.99	0.0	0.90
<b>DKTm+BN</b>	0.0	0.95	0.0	0.97	0.0	0.89	0.19	0.88	0.0	0.97	0.12	0.85	0.0	0.99	0.1	0.89

Pour les connaissances faciles à maîtriser (par exemple le MPP), tous les modèles prédisent toujours que les apprenants donneront des réponses correctes (le score F1 des réponses incorrectes est 0 pour le DKT et presque 0 pour les deux autres modèles) même après l'application du masque sur la fonction de perte. Ceci est dû principalement au fait que l'ensemble des données comporte très peu de réponses incorrectes (6 au total) pour cet élément de connaissance.

**Roger NKAMBOU, Ange TATO, Janie BRISSON, Serge ROBERT,  
Maxime SAINTE-MARIE**

Nous avons testé les modèles avec des valeurs élevées du paramètre de régularisation  $\lambda$  (initialement fixé à 0,10) et nous avons obtenu des valeurs de f1score égales à environ 0,6 pour les bonnes réponses et à environ 0,7 pour les mauvaises réponses. Ce résultat peut être satisfaisant dans d'autres contextes, mais dans le contexte du raisonnement logique où il est établi que le MPP est une connaissance qui est toujours bien maîtrisée, obtenir 0.6 comme f1score pour la prédiction de réponses correctes n'est pas acceptable. C'est pourquoi nous avons gardé  $\lambda$  à 0,10. Cependant, la solution reste valable pour les données où le rapport « *nombre de bonnes réponses/nombre de questions répondues* » ou « *nombre de mauvaises réponses/nombre de questions répondues* » sur une compétence n'est pas trop bas (comme dans ce cas) et est inférieur à 0,5. Pour les connaissances difficiles à maîtriser (connaissances plafonds), il existe une différence significative entre le DKT et les autres modèles. Le DKT de base n'a pas été en mesure de prédire les bonnes réponses sur les connaissances difficiles à maîtriser. Ce comportement ne peut pas être toléré car le traçage des connaissances échouera pour les apprenants qui maîtrisent cette connaissance. Il est donc important de s'assurer que le modèle final soit précis pour toutes les connaissances. Sur les connaissances planchers/plafonds, on constate que le DKTm et le DKTm+BN se comportent mieux que le DKT, ce qui signifie que les changements que nous avons apportés au DKT original n'affectent pas sa capacité prédictive normale.

Ces résultats montrent que le DKT n'est pas en mesure de faire un traçage avec précision des connaissances planchers et plafonds, comparativement au DKTm et au DKTm+BN. Au cours de nos expériences, nous avons remarqué que, pour les connaissances très faciles à maîtriser, les 3 modèles n'ont pas été en mesure de tracer les réponses incorrectes, en raison du fait que le rapport « *nombre de réponses incorrectes/nombre total des réponses aux questions* » était très faible. Cependant, avec le modèle hybride, les résultats ont été notables. Nous avons fixé les valeurs des paramètres de régularisation, mais pour les travaux futurs nous prévoyons de faire une recherche par quadrillage (*grid search*) pour trouver les meilleures valeurs. Nous sommes également conscients que le manque de données pourrait biaiser les résultats obtenus puisque les architectures d'apprentissage profond fonctionnent mieux sur des ensembles de données plus importants. Cependant, le problème des connaissances plancher/plafond peut toujours survenir même avec un grand ensemble de données. Ainsi, nous pensons que les solutions que nous avons proposées peuvent fonctionner sur des ensembles de données plus importants.

## **5. Le module Tuteur de Muse-Logique**

### **5.1. Objectif principal**

L'objectif principal dans Muse-Logique est d'aider l'apprenant à devenir un bon raisonneur logique quelles que soient les situations de raisonnement. Muse-Logique Tuteur doit être en mesure de diagnostiquer la procédure mentale de raisonnement de l'apprenant pour détecter les erreurs de raisonnement et l'aider à les corriger. Pour ce faire, en collaboration avec les experts, nous avons précisé les interventions du tuteur lorsque des erreurs sont détectées ainsi que la façon dont le tuteur peut apporter son aide pour les corriger. Par exemple, dans un problème simple de syllogisme en situation concrète causale avec beaucoup d'alternatives (situation DCCMA), si l'apprenant décide de conclure à un problème invalide, le système doit vérifier si la réponse de l'apprenant vient du fait qu'il n'a pas examiné d'autres alternatives possibles de la situation mentionnée dans la prémisse. Si tel est le cas, le tuteur lui recommandera alors de penser à l'existence d'alternatives. Si l'apprenant refuse de conclure à un problème valide, le tuteur lui demandera la raison de cette réserve. Si sa réponse montre qu'il doute de la véracité de la prémisse majeure, le système lui rappellera l'importance de toujours considérer celle-ci comme pouvant être vraie.

Nous avons implémenté une base de telles règles d'intervention fournies par les experts logiciens pour chaque erreur possible suivant la situation (type de contenu) dans laquelle le raisonnement est opéré. En outre, l'un de nos objectifs était de mettre en place un soutien métacognitif à l'apprenant. Nous avons examiné attentivement les métastructures logiques (pour la logique propositionnelle classique) comme décrit dans les sections précédentes et avons analysé la façon dont elles seront utilisées pour fournir une rétroaction visuelle à l'apprenant, afin qu'il ou elle soit consciente de ses propres erreurs de raisonnement.

### **5.2. Implémentation et intégration des composants validés**

Muse-Logique offre quatre niveaux d'activités d'apprentissage pour l'apprenant, organisés en trois groupes de services d'apprentissage :

- le service d'exploration du domaine qui utilise l'ontologie du domaine ;
- le service des exercices permettant à l'apprenant de résoudre une variété de problèmes de raisonnement incluant les syllogismes et les

polysyllogismes, mais aussi des exercices généraux sur les concepts du raisonnement y compris la conversion de formules logiques dans des situations de raisonnement, la vérification de formules bien formées, la construction de table de vérité, etc. Ce service intègre un générateur automatique de problèmes (voir plus loin) ;

- l'espace de visualisation métacognitif du raisonnement qui offre entre autres la possibilité d'explorer l'état du réseau bayésien, et d'illustrer les erreurs par les métastructures de la logique (treillis de Boole et groupes de Klein pour la logique propositionnelle), complété par un sous-espace d'autodiagnostic offert à l'apprenant et lui permettant de simuler son raisonnement à travers des structures sémantiques comme une table de vérité.

Selon cette dernière perspective, la figure 8 présente un cas d'erreur de raisonnement, avec l'intervention du Tuteur visant à ce que l'apprenant réalise lui-même l'erreur qu'il a commise, et ce par une simulation interactive de son raisonnement à travers une table de vérité.

The screenshot shows the Muse-Logique web application interface. At the top, there is a navigation bar with 'Accueil', 'Services', 'Inscription', 'Qui sommes nous', and 'Plus...'. A user profile 'ariasetato' and a 'Déconnexion' button are visible on the right. Below the navigation bar, there are tabs for 'Exercices', 'Exploration du domaine', and 'Visualisation cognitive'. The main content area is divided into two sections. The left section, titled 'Zone de travail', contains the text 'La forme propositionnelle est:' followed by a text input field containing '((p-q)&q)-p'. Below this is the instruction '1. Complete cette table de vérité:' followed by a truth table with 7 columns and 4 rows. The right section, titled 'Feedback du tuteur', contains the text 'L'énoncé du problème était:' followed by a 'Réponse:' field with 'Si un renard dort, alors il aura une trouille.' and an 'Information:' field with 'Un renard a une trouille.'. Below this is the text 'Tu as répondu " Le renard a dormi"'. A video player icon is visible on the right. At the bottom of the interface, there are buttons for 'Valider', 'Suivant', and 'Quitter', along with navigation links for 'Instructions', 'Aide', 'Leçons', and 'Résultats'.

(p	>	q)	&	q)	>	P
V	V	V	V	V	V	V
V	F	F	F	F	V	V
F	V	V	V	V	F	F
F	V	F	F	F	V	F

**Figure 8 • Exemple d'erreur de raisonnement d'un apprenant avec feedback du Tuteur**

### **5.3. Gérer les erreurs de raisonnement dans Muse-Logique**

Les erreurs systématiques en logique classique et en logique probabiliste sont identifiées dans la littérature ; les erreurs dans d'autres systèmes seront recueillies par des expériences. Dans Muse-Logique, les expériences sur les erreurs de raisonnement sont introduites à travers une variété d'activités de raisonnement, y compris syllogismes et polysyllogismes. Dans certains



d'entre eux, l'apprenant suggère sa réponse en choisissant dans une liste (celle étant valide ou celle se rapportant aux erreurs de type 1), puis le système vérifie (grâce à Muse-Expert) si la réponse suggérée est logiquement valide ou non et fournit des commentaires pertinents après un diagnostic cognitif. Certaines questions offrent un temps de réponse limité de quelques secondes, de sorte que le caractère spontané des réponses de type 1 sera plus facilement déterminé. D'autres invitent les participants à décrire brièvement la procédure utilisée pour répondre, de sorte que cela peut aider pour l'interprétation des résultats et peut révéler des différences de procédure entre les réponses de type 1 et de type 2.

#### **5.4. Génération automatique de problèmes de raisonnement**

Les problèmes de raisonnement de type syllogismes et polysyllogismes sont des cas typiques de résolution de problème en raisonnement logique, car ils nécessitent la capacité de raisonner en conformité avec les règles d'inférence pertinentes pour trouver la solution (parvenir à une conclusion logiquement valide). Un syllogisme catégorique est composé de trois parties, la prémisse majeure, la mineure et la conclusion, par exemple :

- Prémisse majeure : « Je vais choisir la soupe ou je vais choisir la salade » ;
- Prémisse mineure : « Je ne vais pas choisir la soupe » ;
- Conclusion : « Je vais choisir la salade ».

Selon le type de l'opérateur utilisé dans la prémisse majeure, un syllogisme pourrait être disjonctif, implicatif ou d'incompatibilité, soit trois types de raisonnements. L'exemple précédent est un syllogisme disjonctif. Dans ces syllogismes, la prémisse majeure peut toujours être considérée comme un antécédent (A) lié à un conséquent (C) par un opérateur logique. En outre, la prémisse mineure est une déclaration spécifique, qui est soit une affirmation de A ou C, ou la négation de A ou C. Chaque cas correspond à une seule règle d'inférence. Par exemple, la règle d'inférence valide associée à un syllogisme est un *modus ponens ponendo* (MPP) sur le conjonctif conditionnel si la prémisse mineure est l'affirmation de A. Par ailleurs, la conclusion est soit l'affirmation ou la négation de la partie de la prémisse qui n'est pas utilisée dans la mineure. Sur cette base, la prémisse majeure est suffisante pour générer un problème de syllogisme. Donc, notre banque de problèmes contient un stock de prémisses majeures auquel on ajoute des alternatives éventuelles pour A ou C. Pour chaque problème de syllogisme, le système fournit à l'apprenant trois choix

possibles. Les deux premiers choix sont l'affirmation et la négation de la partie de la prémisse majeure qui n'est pas présente dans la mineure. Le troisième choix est « Vous ne pouvez pas conclure ». Un de ces trois choix de réponses est toujours l'erreur spontanée typique de Type 1. Notre algorithme est capable de détecter automatiquement ce choix ainsi que la bonne réponse. Le système tient compte de l'évolution de l'étudiant dans ses apprentissages (état du réseau bayésien) et de la stratégie pédagogique en jeu (règles du tuteur sur l'enchaînement des questions) pour la génération du prochain problème. Les problèmes générés sont situés dans une des six classes de situations identifiées du Tableau 1.

## **6. Conclusion**

Le travail présenté dans cet article est issu d'une initiative multidisciplinaire qui nous a permis de nous inscrire dans une perspective de conception participative ayant mené à la co-construction d'un ensemble de composants validés par les experts du raisonnement logique et de la psychologie du raisonnement avant leur implémentation informatique. Cet article avait pour but de partager cette expérience avec le lecteur. Nous y avons présenté les éléments qui nous ont motivés à proposer un tel système, puis la démarche entreprise pour définir et expliciter les éléments de référence que nous avons utilisés. Nous avons donné un aperçu des différents modules et de leurs composants, en accordant une attention particulière à la construction d'un modèle apprenant prédictif hybride, fondé à la fois sur les données et sur les connaissances expertes. Nous avons décrit les différentes techniques d'apprentissage machine qui ont été développées dans cette perspective. Une première expérimentation a permis de collecter des données pour la construction de ce modèle prédictif comportant une composante psychométrique utilisée pour l'initialisation et la validation du réseau bayésien, un processus de traçage de connaissance bayésienne et un modèle d'apprentissage profond hybride qui lui confère une précision inférentielle améliorée. Les fonctionnalités du système MUSE-Logique ont ensuite été décrites et illustrées.

Muse-Logique est en cours de déploiement dans le cadre d'un cours de logique offert aux étudiants de première année à l'Université de Québec à Montréal. Ce déploiement nous permettra de collecter plus de données empiriques afin de mieux mesurer l'apport du système en termes de gain d'apprentissage, mais aussi, de disposer d'une plus grande quantité de données pour l'entraînement du modèle prédictif de l'apprenant.

## RÉFÉRENCES

- Almond, R. G., DiBello, L. V., Moulder, B. et Zapata-Rivera, J. D. (2007). Modeling diagnostic assessments with Bayesian networks. *Journal of Educational Measurement*, 44(4), 341-359.
- Barnes, T. et Stamper, J. (2010). Automatic hint generation for logic proof tutoring using historical data. *Educational Technology & Society*, 13(1), 3-12.
- Beck, J.E. et Chang, K.-M. (2007). Identifiability: A fundamental problem of student modeling. Dans C. Conati, K. McCoy et G. Paliouras (dir.), *Georgios Proceedings of the 11<sup>th</sup> International Conference on User Modeling* (p. 137-146). Springer.
- Brisson, J., de Chantal, P.-L., Lortie-Forgues, H. et Markovits, H. (2014). Belief bias is stronger when reasoning is more difficult. *Thinking & Reasoning* 20(3), 37-41.
- Chen, J., de la Torre, J. et Zhang, Z. (2013). Relative and absolute fit evaluation in cognitive diagnosis modeling. *Journal of Educational Measurement*, 50(2), 123-140.
- Chen, Y., Yang, Z., Xie, Y. et Wang, Z. (2018). Contrastive learning from pairwise measurements. Dans S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi et R. Garnett, *Proceedings of the 32<sup>nd</sup> Conference on Neural Information Processing Systems (NeurIPS 2018)*. Red Hook.
- Cloutier, A. (2016). *Métacognition, raisonnement logique et philosophie pour enfants* (Mémoire de maîtrise, Université du Québec à Montréal, Canada).
- Corbett, A. T. et Anderson, J. R. (1994). Knowledge tracing: Modeling the acquisition of procedural knowledge. *User modeling and user-adapted interaction*, 4(4), 253-278.
- Cosmides, L. (1989). The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition*, 31(3), 187-276.
- de la Torre, J. (2009). DINA model and parameter estimation: A didactic. *Journal of Educational and Behavioral Statistics*, 34(1), 115-30.
- de la Torre, J. (2011). The generalized DINA model framework. *Psychometrika*, 76, 179-199 (2011).
- Dennett D. (2003). *Freedom evolves*. Viking Press.
- Espino, O. et Byrne, R.M.J. (2020). The suppression of inferences from counterfactual conditionals. *Cognitive Science*, 44.
- Evans, J. S. B. T., Newstead, S. E. et Byrne, R. M. J. (1993). *Human Reasoning: The psychology of deduction*. Lawrence Erlbaum.
- Evans, J. S. B. (2002). Logic and human reasoning: an assessment of the deduction paradigm. *Psychological Bulletin* 128(6), 978-996.
- Giroto, V., Gilly, M., Blaye, A. et Light, P. (1989). Children's performance in the selection task: Plausibility and familiarity. *British Journal of Psychology*, 80(1), 79-95.
- Groß, J., Robitzsch, A. et George, A. C. (2016). Cognitive diagnosis models for baseline testing of educational standards in math. *Journal of Applied Statistics*, 43(1), 229-243.
- Guilford, J.P. et Lyons, T.C. (1942). On determining the reliability and significance of a tetrachoric coefficient of correlation. *Psychometrika*, 7(4), 243-249.
- Houdé, O. (2019). *3-System theory of the cognitive brain: A post-piagetian approach to cognitive development*. Routledge.

**Roger NKAMBOU, Ange TATO, Janie BRISSON, Serge ROBERT,  
Maxime SAINTE-MARIE**

Houdé, O. (2000). Inhibition and cognitive development: Object, number, categorization, and reasoning. *Cognitive Development*, 15(1), 63-73.

Kasurinen, J. et Nikula, U. (2009). Estimating programming knowledge with bayesian knowledge tracing. *ACM SIGCSE Bulletin*, 41(3), 313-317.

Khajah, M., Lindsey, R. V. et Mozer, M. C. (2016). How deep is knowledge tracing? Dans T. Barnes, M. Chi et M. Feng (dir.), *Proceedings of the 9<sup>th</sup> International Conference on Educational Data Mining (EDM 2016)* (p. 94-101). ERIC.

Lesta, L. et Yacef, K. (2002). An intelligent teaching-assistant system for Logic. Dans S. Cerri, G. Gouardères et F. Paraguaçu (dir.), *Proceedings of the 6th International Conference ITS 2002* (p.421-431). Springer-Verlag.

Markovits, H. (2014). On the road toward formal reasoning: Reasoning with factual causal and contrary-to-fact causal premises during early adolescence. *Journal of Experimental Child Psychology*, 128, 37-51.

Markovits, H., Brunet, M.-L., Thompson, V. et Brisson, J. (2013). Direct evidence for a dual process model of deductive inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39(4), 1213-1222.

Mercier, H., Boudry, M., Paglieri, F. et Trouche, E. (2017). Natural-born arguers: Teaching how to make the best of our reasoning abilities. *Educational Psychologist*, 52(1), 1-16.

Moshman, D. (2004). From inference to reasoning: The construction of rationality. *Thinking & Reasoning*, 10(2), 221-239.

Nkambou, R., Mizoguchi, R. et Bourdeau, J. (dir.). (2010). *Advances in Intelligent Tutoring Systems*. Springer.

Piech, C., Bassen, J., Huang, J., Ganguli, S., Sahami, M., Guibas, L.-J. et Sohl-Dickstein, J. (2015). Deep knowledge tracing. Dans *Proceedings of the 28<sup>th</sup> International Conference on neural information processing systems (NeurIPS 2015)* (p. 505-513). ACM.

Robert, S. (2005). Categorization, reasoning and memory. Dans H.Cohen, et C. Lefebvre (dir.), *Handbook of Categorization in Cognitive Science* (p. 699-717). Elsevier Science.

Robert, S. et Brisson, J. (2016). The Klein Group, squares of opposition and the explanation of fallacies in reasoning. *Logica Universalis* 10(2-3), 377-389.

Rossi, S. et Van der Henst, J. B. (dir.). (2007). *Psychologies du raisonnement*. DeBoeck.

Rupp, A., Templin, J. et Henson, R. (2010). *Diagnostic measurement: Theory, methods and applications*. Guilford Press.

Stanovich, K. E. (2011). *Rationality and the reflexive mind*. Oxford Univ. Press.

Tato, A. et Nkambou, R. (2019). Some Improvements of Deep Knowledge Tracing. Dans R. Keefer (dir.), *Proceedings of the IEEE 31<sup>st</sup> International Conference on Tools with Artificial Intelligence (ICTAI)* (p. 1460-1464). IEEE Computer Society.

Tato, A. et Nkambou, R. (2020). Improving first-order optimization algorithms (student abstract). *AAAI-20 Student Tracks*, 34(10), 13935-13936.

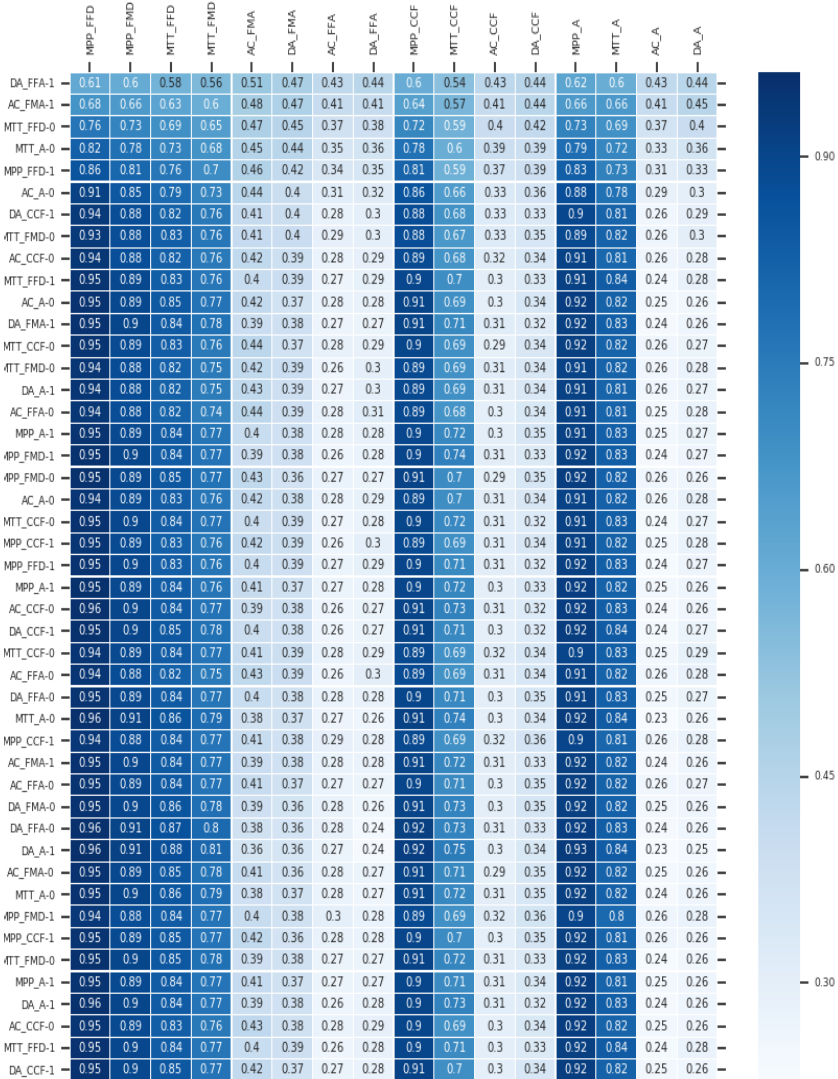
Tchetagni, J., Nkambou, R. et Bourdeau, J. (2007). Explicit Reflection in Prolog-Tutor. *International Journal on Artificial Intelligence in Education*, 17(2), 169-215. <https://telearn.archives-ouvertes.fr/hal-00190040/document>

Varin, C. (2007). *L'Enseignement du raisonnement conditionnel : de la logique aux neurosciences* [mémoire de maîtrise, Université du Québec à Montréal, Canada].

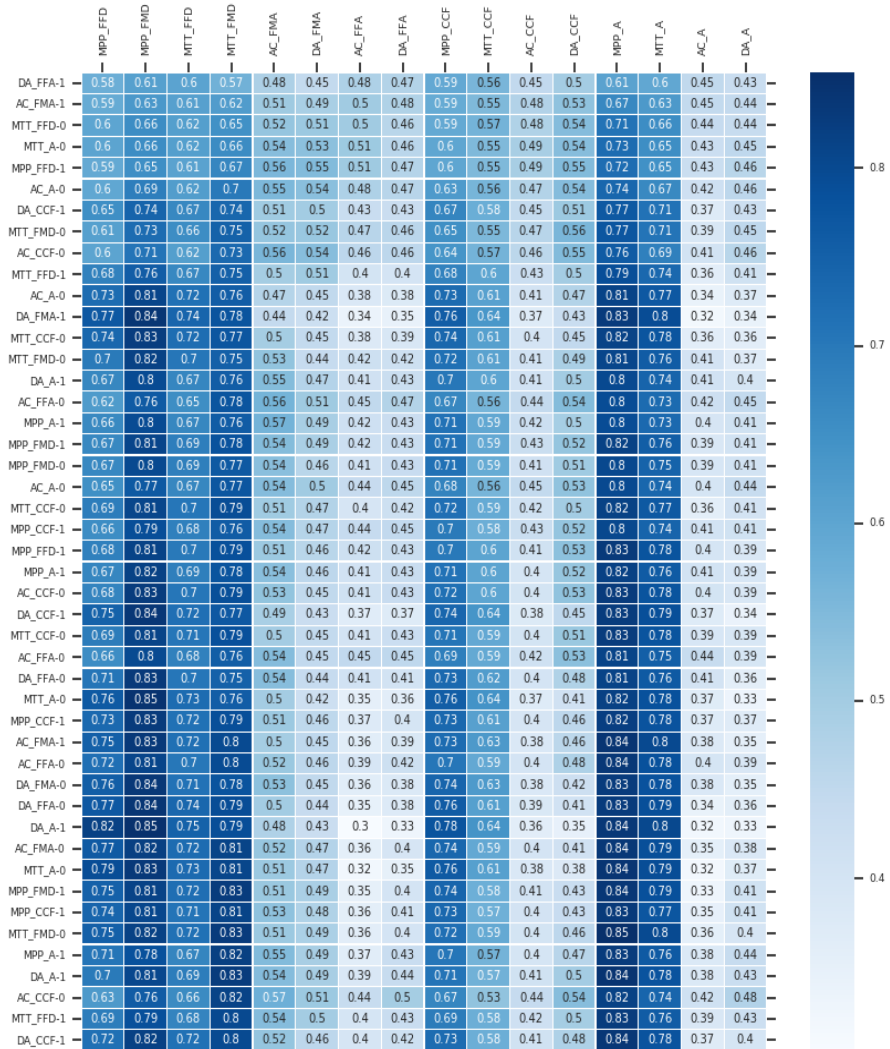
Yudelson, M. V., Koedinger, K. R. et Gordon, G. J. (2013). Individualized Bayesian knowledge tracing models. Dans H. C. Lane, K. Yacef, J. Mostow, et P. Pavlik, *Proceedings of the 16<sup>th</sup> International Conference on Artificial Intelligence in Education (AIED 2013)*(p. 171-180). Springer.

**Roger NKAMBOU, Ange TATO, Janie BRISSON, Serge ROBERT, Maxime SAINTE-MARIE**

**ANNEXE**

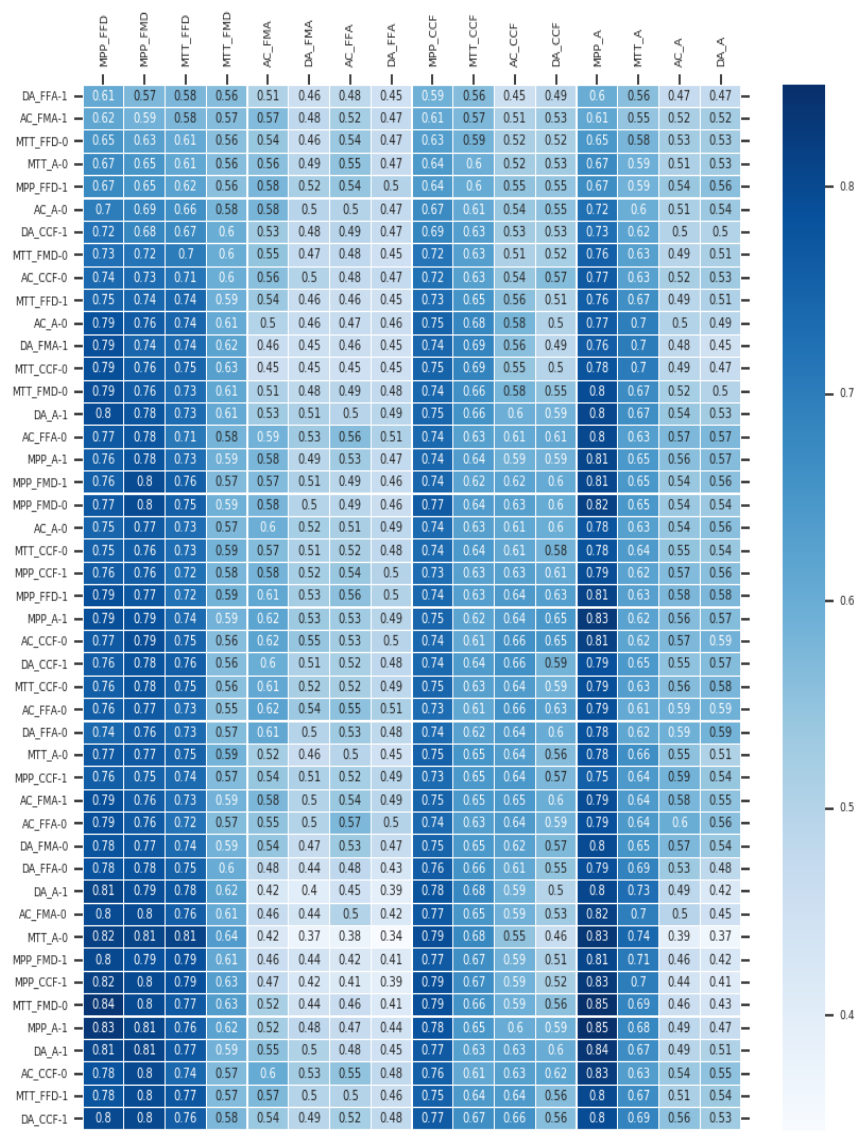


**Figure A1 • Heatmap illustrant la prédiction du DKT original sur le même apprenant**



**Figure A2 • Heatmap illustrant la prédiction du DKTm sur le même apprenant**

# Roger NKAMBOU, Ange TATO, Janie BRISSON, Serge ROBERT, Maxime SAINTE-MARIE



**Figure A3 • Heatmap illustrant la prédiction du DKTm+BN sur le même apprenant**





## Équiper et évaluer un « certificat de compétences cliniques »

► **Olivier CATTEAU** (IRIT, Université de Toulouse), **Anne MAYÈRE** (CERTOP, Université de Toulouse), **Nicolas SAVY** (IMT, Université de Toulouse), **Odile BEYNE-RAUZY** (CHU de Toulouse, Université de Toulouse)

---

---

■ **RÉSUMÉ** • Cet article présente les résultats de l'étude de la mise en place d'un certificat de compétences cliniques à travers la conception et la mise en œuvre d'un environnement informatique. Cet environnement informatique vise, d'une part, à équiper cet examen et, d'autre part, à mener une démarche réflexive le concernant. Les résultats des étudiants à cet examen montrent l'absence de redondance avec ceux des épreuves classantes typiques de ce parcours de formation. En confrontant les retours des étudiants au travers de questionnaires anonymes et l'étude du dispositif de formation lui-même, nous interrogeons l'approche qui le sous-tend concernant les compétences en général et les compétences relationnelles en particulier.

■ **MOTS-CLÉS** • Approche par compétences, évaluation, modèles pédagogiques, collecte de données, traitement de données, ECOS.

■ **ABSTRACT** • *This paper presents the results of the study of the implementation of a clinical aptitude certificate through the design and implementation of an IT environment. This IT environment aims on the one hand to equip this examination, and on the other hand to conduct a reflexive approach concerning it. The students' results in this examination show no redundancy with those of the ranking tests typical of this training program. By comparing student feedback through anonymous questionnaires and the study of the training system itself, we question the underlying approach to skills in general and relational skills in particular.*

■ **KEYWORDS** • *Competency approach, assessment, learning model, data collection, data mining, OSCE.*

## **1. Introduction**

Un certificat de compétences cliniques a été mis en place depuis 2016 à l'issue du deuxième cycle des études médicales en France (année bac+6). Ce certificat représente une innovation relative pour ce domaine, dans la mesure où il suppose de rompre avec une logique dominante de concours et d'évaluation de connaissances via des questionnaires à choix multiple (QCM), pour évaluer des compétences, en l'occurrence cliniques. Cet article prend appui sur les résultats d'une démarche initiée dans le cadre d'un programme de recherche interdisciplinaire financé par l'Institut de Recherche en Santé Publique (IReSP)<sup>1</sup>. À travers la conception et la mise en œuvre d'un environnement informatique, il s'agit d'une part d'équiper le dispositif d'évaluation requis pour ce certificat et, d'autre part, de contribuer à un retour réflexif sur le dispositif.

Nous exposerons tout d'abord le contexte d'introduction de ce nouveau certificat, ainsi que les conditions de sa mise en œuvre à l'Unité de formation et de Recherche (UFR) de Médecine de Toulouse Purpan. À travers l'analyse des résultats de cet examen, nous montrerons que ce certificat se distingue des autres dispositifs d'évaluation, ce qui conduit à considérer qu'il évalue des dispositions ou des savoirs autres que ceux que mesurent les épreuves classantes.

Sur la base de l'analyse des résultats d'examen, qui permet d'en préciser les modalités et attendus, nous proposons une prise de recul : qu'en est-il de ce qui est ainsi évalué ? Tant les retours des étudiants via des questionnaires anonymes, que la mise en discussion des résultats intermédiaires avec ceux d'autres recherches, invitent à interroger les approches pédagogiques qui sous-tendent le dispositif et les acceptions associées des compétences cliniques.

Enfin, nous concluons en proposant des perspectives de recherche.

---

<sup>1</sup> Voir les remerciements en fin d'article.

## **2. Conditions d'élaboration et mise en œuvre d'un dispositif d'évaluation**

### **2.1. Contexte de développement d'un certificat de compétences cliniques**

Dans le cadre de la réforme des études médicales, la définition d'un référentiel de compétences génériques dans le Bulletin officiel du 16 mai 2013 a rendu nécessaire une transformation pédagogique pour mettre en place un processus d'acquisition, de développement et de certification des compétences cliniques en médecine (2e cycle). Le certificat de compétences cliniques (CCC) a été mis en place pour en assurer la validation (MESRI, 2013) ; les principales compétences attendues sont celles de : « Clinicien », « Communicateur », « Acteur de Santé Publique », « Scientifique », « Responsable aux plans éthique et déontologique », « Réflexif » (les extraits du BO sont présentés en annexe).

Beaucoup d'auteurs se sont penchés sur la notion de compétence. Nous utiliserons dans cet article la définition de Tardif (2006) : « *Un savoir-agir complexe prenant appui sur la mobilisation et la combinaison efficaces d'une variété de ressources internes et externes à l'intérieur d'une famille de situations.* » Cette définition souligne le caractère complexe et situé des compétences. Or l'inscription de cette démarche par compétences dans les études médicales vient en tension avec une logique de concours, qui suppose des critères objectifs et « neutres » d'évaluation pour l'interclassement des différents candidats. Cette logique du concours a progressivement transformé les études médicales, en faisant du QCM la modalité prédominante d'évaluation, et l'organisation des enseignements. Le nouveau certificat de compétences cliniques se trouve ainsi « pris en tenaille » entre l'objectif d'évaluer des savoirs complexes, interactionnels et situés, et l'exigence de critères d'évaluation objectivables et standardisables.

Nous proposons dans le cadre de cet article de questionner les logiques en tension dans un tel dispositif d'évaluation. Nous verrons qu'une forme première de résolution de ces tensions a consisté à privilégier une approche behavioriste, focalisée sur des comportements observables et mesurables (Boutin, 2004). Nous poserons ensuite des jalons vers d'autres approches susceptibles de mieux évaluer les compétences mobilisées pour affronter des situations qui réclament une réflexion sur et dans l'action (Boutin, 2000).

Le dispositif d'évaluation retenu permet de placer les étudiants en situation professionnelle et constitue une épreuve appelée « Examen Clinique Objectif Structuré » (ECOS), largement utilisée dans les pays anglo-saxons pour la certification des médecins (Khanz *et al.*, 2013a). Les étudiants vont successivement être interrogés en situation clinique au sein d'une station thématique pendant un temps limité de 7 minutes, l'épreuve comptant au total 7 stations. Le choix de ces stations correspond à des compétences dont l'acquisition est indispensable en fin de deuxième cycle des études médicales (6e année). Avant de décrire les stations, deux éléments doivent être précisés :

- la définition d'un patient-standardisé, qui est une personne formée à reproduire de façon standardisée une situation clinique élaborée spécifiquement par l'équipe pédagogique pour évaluer certaines compétences, cette personne est habituellement un comédien ;
- l'utilisation des mannequins qui permettent de reproduire des anomalies de l'examen clinique, par exemple une anomalie de l'auscultation cardiaque ou encore une tumeur du sein, dans les stations dévolues à l'évaluation de l'examen clinique.

Les deux premières stations de l'épreuve d'ECOS sont les stations «électrocardiogramme (ECG)» et «Thérapeutique». Elles abordent des situations médicales d'urgence. La station suivante, «Dermatologie» devenue ensuite «Iconographie», utilise un document (photo, radiographie) à analyser dans une situation clinique. La station « Annonce » et la station « Interrogatoire » utilisent un patient-standardisé et évaluent respectivement les aptitudes relationnelles de l'étudiant à délivrer une information au patient ou à solliciter des informations dans le cadre d'une démarche de diagnostic. La station « Clinique » utilise un mannequin permettant d'évaluer l'étudiant.e en situation d'examiner un patient. La station « Examens complémentaires » évalue la pertinence d'une demande d'examen complémentaire dans une situation clinique et son interprétation. Enfin la station « Synthèse » permet à l'étudiant.e d'argumenter l'ensemble de la démarche concernant les stations « Interrogatoire », « Clinique » et « Examens complémentaires » ; seule cette station comporte deux examinateurs.

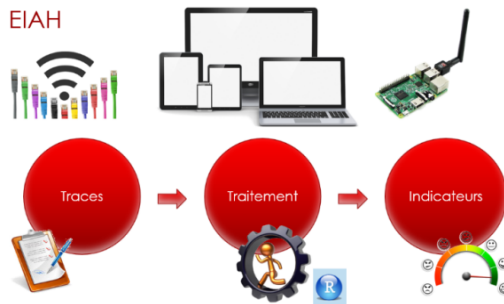
À chaque station, l'examineur utilise des grilles critériées qui visent à favoriser l'homogénéité de l'évaluation (Khanz *et al.*, 2013b).

En parallèle à la mise en place du nouveau dispositif d'évaluation, le dispositif pédagogique global a été modifié afin de permettre la formation des étudiants à ces nouvelles modalités d'évaluation, incluant en particulier des séances de simulation avec des patients-standardisés telles que pratiquées dans deux stations de l'examen, ou encore la formation à l'examen clinique ou à certains gestes sur mannequin pendant l'ensemble de leur cursus.

## 2.2. Méthodologie

Pour la mise en œuvre de cet examen, un environnement informatique a été mis en place. Il est représenté sur la figure 1. Son rôle est de recueillir les données d'évaluations, d'automatiser et d'accélérer leurs traitements, pour faire une remontée des résultats plus rapidement.

L'environnement, essentiellement composé à ce stade de grilles d'évaluation, permet de collecter sous forme de traces brutes un ensemble d'éléments observés. Côté *front office*, le recueil de données se fait au travers d'une application *web responsive* permettant aux évaluateurs de saisir leurs résultats.

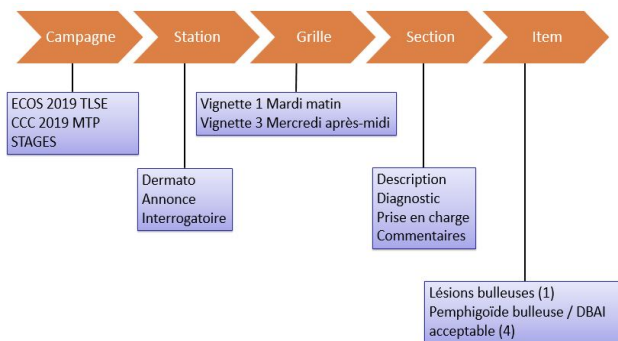


**Figure 1 • Environnement informatique mis en place**

Le certificat de compétences cliniques est organisé dans les locaux de l'Institut Toulousain de SIMulation en Santé (ITSIMS) qui dispose d'une infrastructure hétérogène. Les utilisateurs peuvent être amenés à utiliser l'environnement informatique aussi bien à partir de postes fixes reliés au réseau filaire, que de PC portables ou de tablettes reliés au WiFi. Quelques salles d'examen ne disposent ni de connexion filaire ni d'une qualité de réception suffisante du WiFi, il a fallu alors installer des systèmes autonomes : bornes WiFi avec portail captif vers une version hors ligne de l'environnement informatique. Les données collectées ont été transférées

toutes les demi-journées sur le serveur central. Les étudiants ont parfois dû utiliser leur téléphone portable avec connexion 4G pour répondre au questionnaire anonyme de fin d'examen.

Le système permet de gérer plusieurs campagnes d'évaluation (une par année, cf. figure 2). À chaque campagne sont rattachées les 7 stations. Pour chaque station, entre 16 et 30 grilles d'évaluation sont créées pour éviter la communication des sujets entre étudiants. Il est possible de créer des grilles spécifiques pour la gestion de questionnaires anonymes. Chaque grille est découpée en 3 ou 4 sections contenant les items à évaluer.



**Figure 2 • Représentation synthétique des données**

À chaque campagne on associe une cohorte d'étudiants. Pour chaque station, le passage de chaque étudiant sur une des grilles est défini à l'avance en tenant compte des parcours (2 étudiants différents peuvent en effet être examinés au même moment par des jurys différents).

L'examineur n'a plus qu'à préciser (1) la station sur laquelle il travaille et (2) l'étudiant qui se présente à lui pour pouvoir compléter la grille d'évaluation (cf. figure 3).

ECOS Février 2019 Toulouse Purpan Clinique 03 - Mardi 12/02 après-midi - Vignette 4 / CELINE

**ATTENTION: n'oubliez pas d'enregistrer en bas de formulaire (quand tous les items obligatoires sont remplis)**

**Sujet**

Résumé du dossier : Patiente 29 ans pas d'antécédent, allergie à l'oeuf, exanthème fébrile et pneumopathie base gauche au décours épisode ORL.

Situation clinique : 39°5, FC 120/min tachycardie, TA 12/8, FR 28, saturation 90% AA, photo exanthème morbilliforme descendant et énanthème conjonctival, pulmonaire foyer de crépitations base gauche et rouchi bilatéraux.

« SAVOIR FAIRE » TRONC COMMUN sur 10 points

Sait prendre le pouls radial	Non fait +/-	Fait
Ausculte correctement le cœur foyer aortique principal (2EIC D)	Non fait +/-	Fait
Ausculte correctement le cœur foyer mitral (5EIC G ligne médio-clav)	Non fait +/-	Fait
Sait mesurer la fréquence respiratoire	Non fait +/-	Fait
Demande la mesure de la saturation en oxygène	Non fait +/-	Fait

Mardi 12/02 après-midi - Vignette 4 - Afficher la liste complète

CELINE ALIXIA SUHEDA FLORENCE MARIE MATHIEU  
 JEREMY DEBORAH GALLA MARGAUX  
 CLEMENCE CHLOE CINDY ARTHUR BERTRAND LOIC  
 AUGUSTA RENAUD MARIE ADELE

**Figure 3 • Évaluation des étudiants**

Chaque clic sur une case d'évaluation fait l'objet d'une trace brute envoyée au *back-office*. Une fois collectées, les traces sont modélisées dans le *back-office*, puis des algorithmes de calcul sont mis en place pour produire des indicateurs (Djouad, 2013).

Les principaux indicateurs à destination des responsables de formation concernent la réussite à l'examen, le taux de satisfaction des enseignants, le taux de satisfaction des étudiants, le taux de difficulté perçue, mais également d'autres indicateurs qui prennent appui sur une analyse statistique et docimologique des grilles d'évaluation et des résultats associés (Dessus, 2019). Une des finalités est de vérifier la pertinence du dispositif d'évaluation et de l'améliorer pour les futures sessions. La vérification de la pertinence du dispositif s'effectuera par une analyse descriptive des résultats (mesures de position et de dispersion des différents résultats) enrichie par une analyse de la corrélation avec d'autres méthodes d'évaluation (Lecoutre, 2016) telles que les épreuves théoriques de 6e année et les épreuves classantes nationales (ECN). Enfin l'analyse peut être affinée au moyen de techniques d'analyse en composantes principales (Husson *et al.*, 2016). En effet une faible chute d'inertie caractérise une isotropie des

résultats pour les différentes composantes de l'évaluation et par conséquent une homogénéité globale des résultats.

Les principaux indicateurs à destination des étudiants relèvent de la rétro-information : réussite à l'examen, commentaires des évaluateurs, évaluation et commentaires des personnes assurant le rôle de patients-standardisés. Lorsque des indicateurs nécessitent des calculs statistiques (par exemple, coefficient de corrélation), le *back office* sollicite l'application R version 3.5.0 (2018-04-23).

Le certificat de compétences cliniques n'a pas à ce stade finalité à participer à l'interclassement des étudiants, mais à vérifier la maîtrise de compétences associées à la pratique clinique. Il ne s'agit pas de reproduire la logique « concours » de la Première Année Commune des Études de Santé (PACES) ou des ECN en favorisant des critères discriminants ; la priorité est donnée à l'homogénéité des grilles d'évaluation.

L'environnement informatique constitué permet par ailleurs de recueillir de façon anonyme les retours tant des enseignants que des étudiants. La réflexivité ainsi proposée aux étudiants est un appui important pour accompagner l'apprentissage et questionner plus avant les compétences attendues et évaluées.

### **2.3. Expérimentation**

À l'UFR de Médecine de Toulouse Purpan, le CCC a donné lieu à une nouvelle modalité d'évaluation sous forme d'un ECOS à destination des étudiants de 6e année.

L'examen a concerné 141 étudiants en janvier 2016, 134 étudiants en février 2017, 138 en 2018 et 151 en 2019. En 2016 et 2017, les étudiants sont passés successivement par les 7 stations suivantes : « ECG, Dermatologie, Thérapeutique, Annonce, Interrogatoire, Examen clinique, Synthèse ». Ces stations formalisent des étapes standards de la pratique clinique. Il s'agit en particulier d'évaluer la capacité à mettre en œuvre un examen clinique et à formuler des hypothèses de diagnostic et/ou de prise en charge. Un lien est établi entre plusieurs stations pour envisager un même cas à différentes étapes de la démarche clinique.

L'ensemble des données collectées et analysées est résumé dans le Tableau 1, qui schématise également les évolutions du dispositif.



**Tableau 1 • Effectifs collectés et analysés de 2016 à 2019**

	2016	2017	2018	2019
Station ECG	141	134		
Station ECG/Thérapeutique			135	151
Station Dermatologie/Iconographie	141	134	135	151
Station Thérapeutique	141	134		
Station Annonce	141	134	135	151
Station Annonce (retour des acteurs)		134	101	
Station Interrogatoire	141	134	135	151
Station Interrogatoire (retour des acteurs)		125	94	
Station Clinique	141	134	135	
Station Examen complémentaire			135	151
Station Synthèse	141	134	135	151
Questionnaire anonyme post-évaluation à destination des étudiants		134	128	140
Entretiens post-évaluation avec les étudiants			11	17
Questionnaire post-évaluation à destination des évaluateurs		30		35
Résultats aux épreuves théoriques de 6e année	131	128	124	151
Résultats des ECN	137	128	123	144

Fusion des stations « ECG » et « Thérapeutique »

Création de la station examen complémentaire

L'environnement informatique n'a été mis en place qu'en 2017 mais une analyse statistique a pu être effectuée a posteriori sur les résultats de 2016. La logistique de l'examen, organisé sur 4 demi-journées, est assez lourde avec environ 25 examinateurs mobilisés, certains participant à plusieurs demi-journées. Chaque étudiant doit passer 7 minutes par station. Le changement de station se fait au coup de sifflet. Chaque station dispose de plusieurs « vignettes » ou sujets d'examens pour éviter la communication entre les étudiants qui ont terminé leur parcours et ceux qui ne sont pas encore passés.

En 2018, le contenu de certaines stations a été modifié pour tenir compte des retours des enseignants: les thématiques « ECG » et « Thérapeutique » ont été regroupées en une seule station. De plus, une nouvelle station « Examen complémentaire » a été introduite à la suite de

l'examen clinique. C'est une autre caractéristique de ce dispositif que d'être réflexif et de supposer un environnement informatique qui équipe cette réflexivité et qui soit aisément évolutif tout en assurant des formes de comparabilité des résultats.

Les capacités mises en œuvre dans les interactions médecin / patient sont plus particulièrement considérées au cours de deux stations faisant intervenir des comédiens comme patients-standardisés (« Annonce » et « Interrogatoire »).

Sur chaque station les examinateurs disposent d'une grille d'évaluation informatisée. Les grilles des stations « ECG », « Dermatologie » et « Synthèse » sont composées de 10 à 20 critères évalués entre 1 et 6 points. Les autres grilles sont composées de 10 à 30 éléments pour lesquels l'évaluateur indique s'ils ont été effectués complètement, partiellement ou pas du tout. Le rythme de passage contraint le temps passé par les enseignants à évaluer. Il suppose un environnement informatique très robuste et soulève des défis quant à sa conception et son évolution pour combiner efficacité et pertinence.

Un travail exploratoire a été mené par les responsables pédagogiques afin de tenter de mettre en correspondance, d'une part, les éléments d'évaluation et, d'autre part, les compétences génériques définies dans l'annexe du BO pour décrire les compétences génériques dont l'acquisition est nécessaire pour l'obtention du certificat : Clinicien, Communicateur, Acteur de Santé Publique, Scientifique, Responsable aux plans éthique et déontologique, Réflexif (MESRI, 2013).

En fin de parcours, l'avis des étudiants et des enseignants a été collecté, sur une base volontaire et anonyme, au travers de questionnaires qui leur étaient dédiés. Par ailleurs, des entretiens semi-directifs ont été menés avec des étudiants, sur une base volontaire, concernant leur formation à la relation médecin-patient, avec des questions plus spécifiques relatives à cet examen. Cette base d'informations est centrale pour la recherche ici menée dans la mesure où elle permet de recueillir des avis indépendants du contexte de l'examen. Ce dernier est en effet achevé au moment où il est proposé aux étudiants de prendre quelques instants pour répondre, ce que la plupart d'entre eux font bien volontiers (100 % de participation en 2017, 95 % en 2018, 93 % en 2019), ce moment constituant une forme de sas de sortie d'examen.

## 2.4. Résultats

L'environnement informatique a permis de diffuser aux candidats les résultats à l'examen en 2017, 2018 et 2019 beaucoup plus rapidement qu'en 2016, en évitant un traitement manuel et potentiellement source d'erreurs.

Les questionnaires post-évaluation montrent un intérêt pour la démarche, tant de la part des étudiants que des enseignants-examineurs. En 2017, le taux de satisfaction du processus d'évaluation pour les enseignants était de 100 % (100 % également en 2019), il était de 92 % pour les étudiants (98 % en 2018, 94 % en 2019). À 76 % les étudiants ont trouvé que l'examen était adapté par rapport à la durée et au niveau demandé (92 % en 2018, 85 % en 2019). En 2017, la station dite « Annonce » (au cours de laquelle une pathologie doit être annoncée au patient standardisé) était perçue comme étant la plus difficile par les étudiants (64 % de taux de difficulté). Les compétences mises en jeu dans la relation soignant-patient, domaine encore peu enseigné au cours des études médicales, sont plus particulièrement mises à l'épreuve dans de tels contextes. La situation a évolué en 2018, la station « Annonce » ayant été mieux préparée avec la mise en place d'enseignements dédiés. Par ailleurs les situations d'annonce ont été simplifiées, en portant sur des annonces d'examen et non directement de pathologie. Le taux de difficulté perçue par les étudiants a alors diminué à 44 % en 2018, 49 % en 2019 alors que la station « Examen clinique sur mannequin » a été perçue comme étant la plus difficile (44 % de taux de difficulté en 2017, 49 % en 2018, 52 % en 2019). Des éléments d'explication peuvent tenir aux limites techniques des mannequins et au manque d'entraînement des étudiants dans ce contexte.

Le retour sur évaluation était très attendu par les étudiants. Les résultats de chaque station ont été communiqués par groupe de critères. Exemples : Savoir Être, Diagnostic, Prise en Charge, Raisonnement... Les commentaires saisis par les évaluateurs étaient également communiqués. De plus, sur les 2 stations où intervenaient des acteurs jouant le rôle de patient-standardisé (« Annonce » et « Interrogatoire »), une évaluation formative du candidat leur était demandée. Le résultat de cette évaluation était communiqué aux étudiants, mais ne rentrait pas dans le calcul de la note finale pour l'attribution du certificat de compétences cliniques. Cette évaluation pourrait toutefois devenir sommative à l'avenir. En 2017, la corrélation entre l'évaluation formative des acteurs et l'évaluation sommative des enseignants sur le savoir être des étudiants n'était que de 0.39 pour l'Annonce ( $p = 3.4e-6$ ) et de 0.31 pour l'Interrogatoire ( $p = 2.427e-4$ ). Il y a là

un élément intéressant à investiguer plus avant pour expliciter, d'une part, les attendus des enseignants et, d'autre part, ceux des acteurs jouant les patients-standardisés qui ont été formés au feedback.

L'examen n'avait pas de visée sélective, et cela s'est confirmé par la réussite de la quasi-totalité des étudiants, seuls un à deux passant chaque année à la session de rattrapage.

Une analyse statistique plus fine des résultats, résumée dans le Tableau 2, montre que chaque station a son utilité dans le dispositif. Les résultats sont calculés de la manière suivante : le score de chaque station est obtenu en faisant la moyenne pondérée de tous les éléments évalués lors du passage de l'étudiant sur la station. Les résultats sont présentés sur un barème de 20 points par station. Un élément effectué complètement rapporte 1 point, un élément effectué partiellement rapporte un demi-point et un élément non effectué ne rapporte pas de point. Le score global est obtenu sur un barème de 100 points en faisant la moyenne pondérée des 7 stations. Prenons l'exemple de la station Dermato en 2016, le coefficient de corrélation est calculé entre (1) les résultats de cette station et (2) la somme des résultats des six autres stations. La normalité des résultats a été vérifiée. Pour chaque calcul de coefficient de corrélation, la méthode suivante a été appliquée :

- (1) vérification que l'allure de l'histogramme des variables est sous forme de cloche,
- (2) vérification que l'allure du diagramme quantile-quantile est sous forme de droite de pente égale à 1, et
- (3) élimination des valeurs isolées (inférieures à  $Q1 - 1.5(Q3 - Q1)$  et supérieures à  $Q3 + 1.5(Q3 - Q1)$ ).

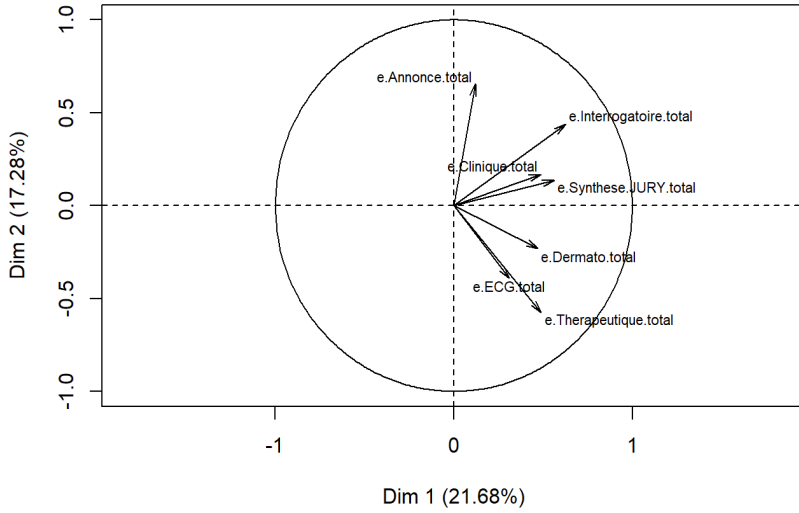
Les coefficients de corrélation calculés entre les résultats des grilles d'une station donnée et les résultats des grilles des autres stations, quand ils sont significatifs, restent positifs et faibles (entre 0.17 et 0.24 en 2016 ; entre 0.22 et 0.36 en 2017 ; entre 0.18 et 0.29 en 2018 ; entre 0.17 et 0.38 en 2019) justifiant ainsi de l'intérêt de chaque station.

**Tableau 2 • Analyse statistique des résultats des différentes stations**

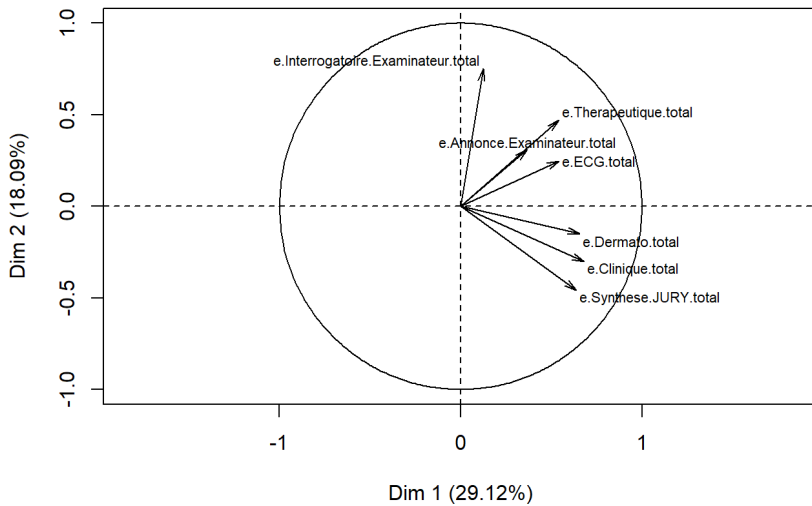
2016									
Station	Min	Q1	Med	Moy	Q3	Max	ET	Cor	p
Global	49.5	68.0	72.8	72.3	77.0	88.2	6.89		
ECG	8.0	14.0	16.0	15.2	18.0	20.0	2.88	non significatif	
Dermato	2.0	11.0	15.0	13.9	17.0	20.0	3.99	0.17	[0.00;0.33] 0.048
Thérapeu	6.0	13.0	15.0	14.6	17.0	20.0	3.26	0.19	[0.02;0.34] 0.029
Annonce	0.0	12.0	15.0	14.1	17.0	20.0	3.80	non significatif	
Interro	5.0	11.5	13.0	13.0	14.5	18.2	2.11	0.24	[0.08;0.39] 0.0039
Clinique	7.5	14.0	15.5	15.1	17.0	20.0	2.43	non significatif	
Synthèse	8.0	13.0	16.0	15.3	18.0	20.0	3.16	0.18	[0.01;0.33] 0.037
2017									
Station	Min	Q1	Med	Moy	Q3	Max	ET	Cor	p
Global	49.5	68.6	73.0	72.5	77.0	85.2	6.89		
ECG	6.5	13.0	16.5	15.6	18.0	20.0	3.34	0.28	[0.11;0.43] 0.0014
Dermato	6.0	12.5	14.5	14.3	16.5	19.5	3.10	0.36	[0.21;0.51] 2.1e-05
Thérapeu	5.0	12.0	14.0	13.7	15.5	19.5	2.66	0.26	[0.10;0.42] 0.0024
Annonce	5.5	13.5	15.0	15.0	16.9	19.5	2.44	non significatif	
Interro	9.8	13.6	14.9	14.8	16.2	18.8	2.03	non significatif	
Clinique	8.0	12.5	14.0	13.9	15.5	17.5	1.95	0.35	[0.20;0.50] 3e-05
Synthèse	4.0	13.0	15.0	14.5	17.0	20.0	3.11	0.22	[0.05;0.38] 0.011
2018									
Station	Min	Q1	Med	Moy	Q3	Max	ET	Cor	p
Global	48.5	67.6	72.5	71.6	76.1	89.5	6.99		
ECGThe	4.0	11.5	14.5	14.0	17.0	19.0	3.32	non significatif	
Dermato	4.0	11.5	13.5	13.5	15.5	19.0	3.04	non significatif	
Annonce	9.5	15.0	17.0	16.4	18.0	20.0	2.39	0.29	[0.12;0.44] 0.00078
Interro	8.8	13.5	15.0	14.8	16.2	19.0	1.96	non significatif	
Clinique	4.5	11.8	13.5	13.3	15.0	17.5	2.61	non significatif	
Compl	2.0	12.0	14.0	13.7	16.0	20.0	3.77	0.18	[0.01;0.34] 0.04
Synthèse	6.5	13.0	15.0	14.8	17.0	19.0	2.70	0.29	[0.12;0.44] 0.00094
2019									
Station	Min	Q1	Med	Moy	Q3	Max	ET	Cor	p
Global	53.5	69.4	75.0	74.3	79.8	89.5	7.24		
ECGThe	5.0	13.0	15.0	14.6	16.8	19.5	2.96	0.21	[0.05;0.36] 0.011
Dermato	5.5	12.5	14.5	14.4	16.5	20.0	3.01	0.31	[0.16;0.45] 9.5e-05
Annonce	8.5	14.5	16.0	16.1	17.5	20.0	2.09	non significatif	
Interro	7.8	13.5	14.8	14.7	15.8	18.5	1.95	0.23	[0.07;0.37] 0.006
Clinique	9.0	12.5	15.0	14.7	17.0	20.0	2.66	0.17	[0.01;0.32] 0.034
Compl	6.0	13.0	15.0	14.7	17.0	20.0	3.42	0.26	[0.10;0.40] 0.0014
Synthèse	5.0	13.0	15.5	15.0	17.0	20.0	3.03	0.38	[0.23;0.51] 1.7e-06

La faible dispersion des résultats montre également l'homogénéité de ces derniers. Une analyse en composantes principales indique une chute d'inertie faible (39 % de l'information sur les 2 premières dimensions en 2016, 47,21 % en 2017, 31,88 % en 2018, 45,39 % en 2019, cf. figure 4). Là encore, on retrouve une homogénéité, il n'y a pas de caractère discriminant entre les stations.

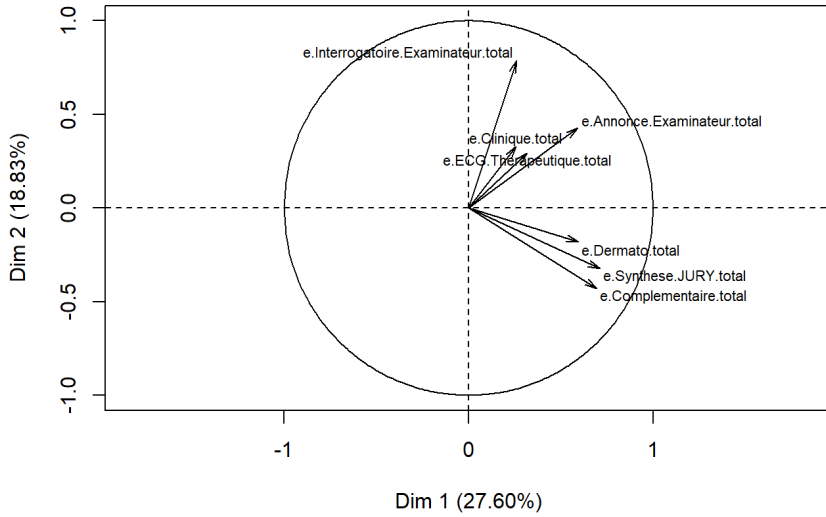
Variables factor map (PCA)



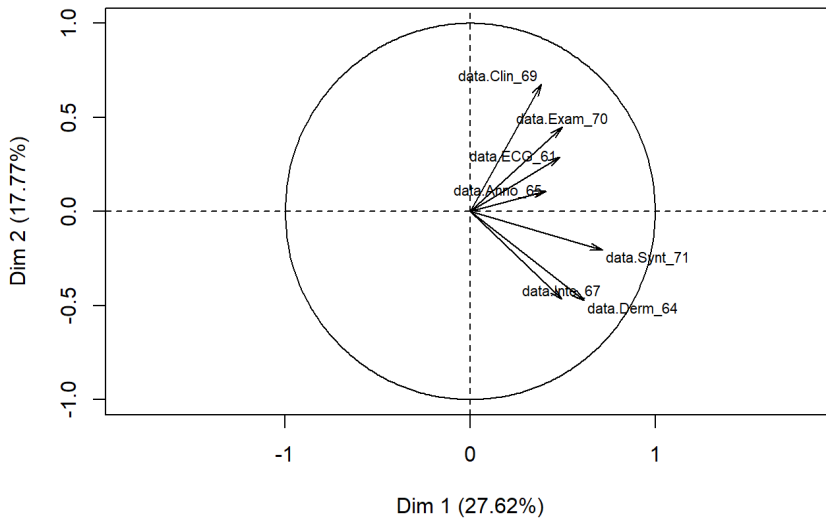
Variables factor map (PCA)



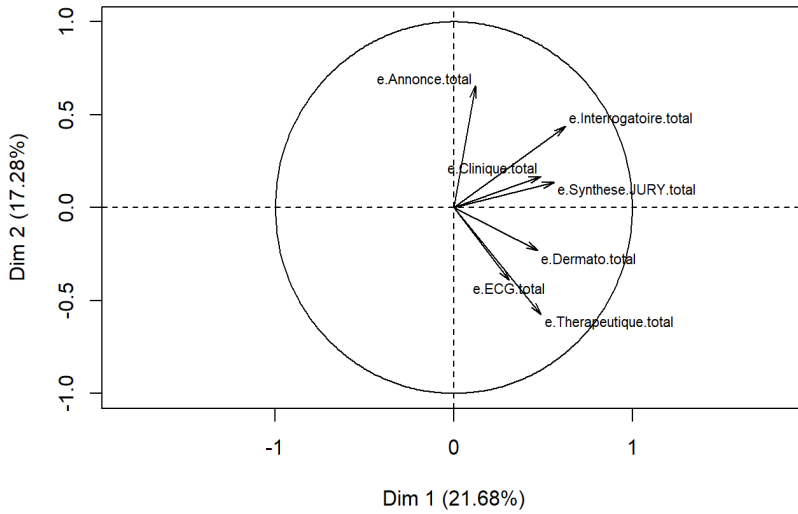
**Variables factor map (PCA)**



**Variables factor map (PCA)**

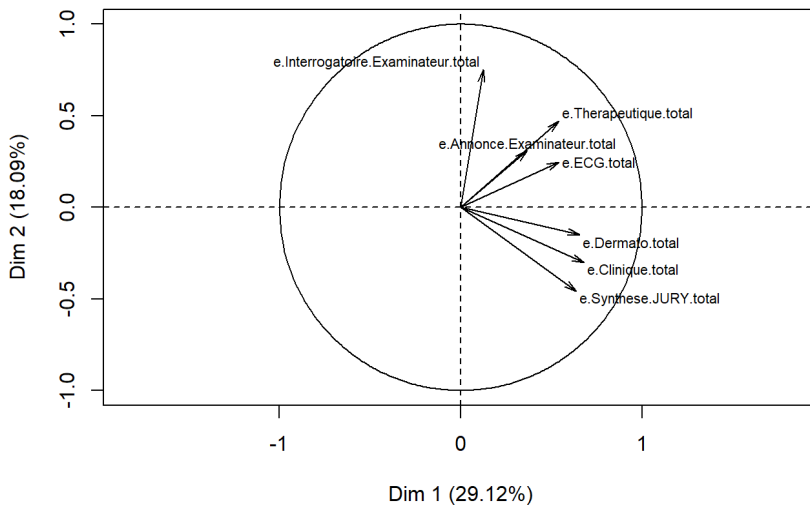


Variables factor map (PCA) 2016

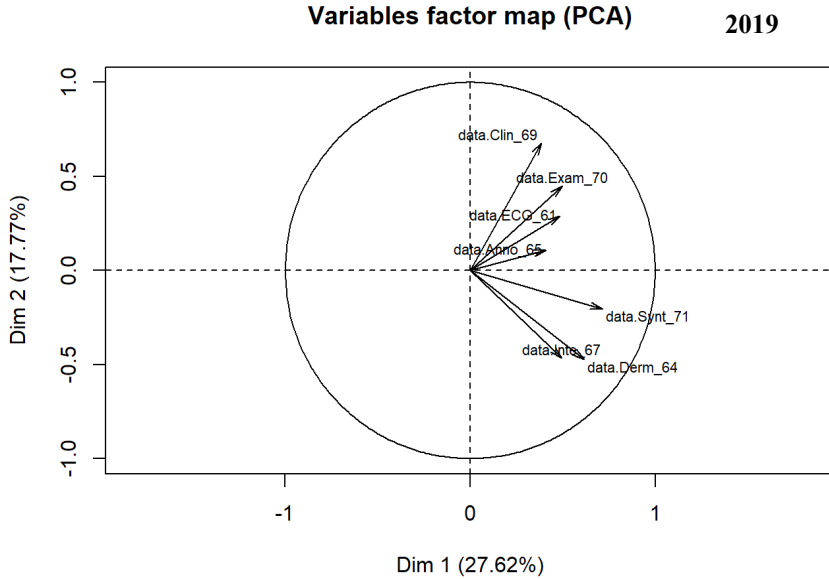
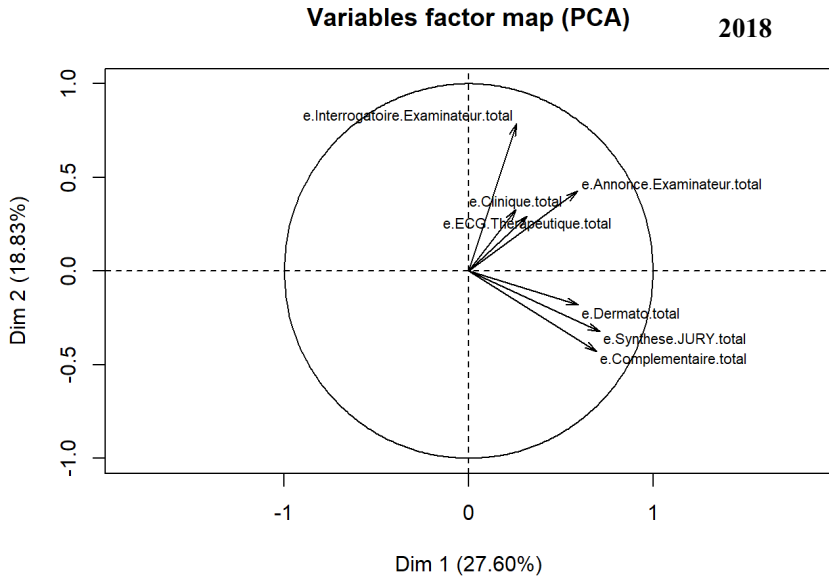


Variables factor map (PCA)

2017

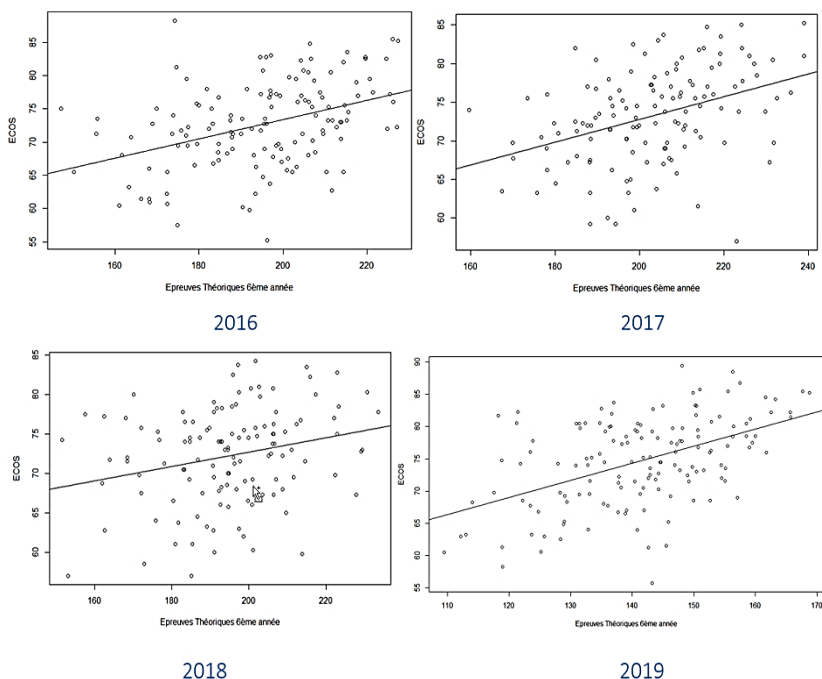






**Figure 4 • Cercles des corrélations issus de l'analyse en composantes principales des données**

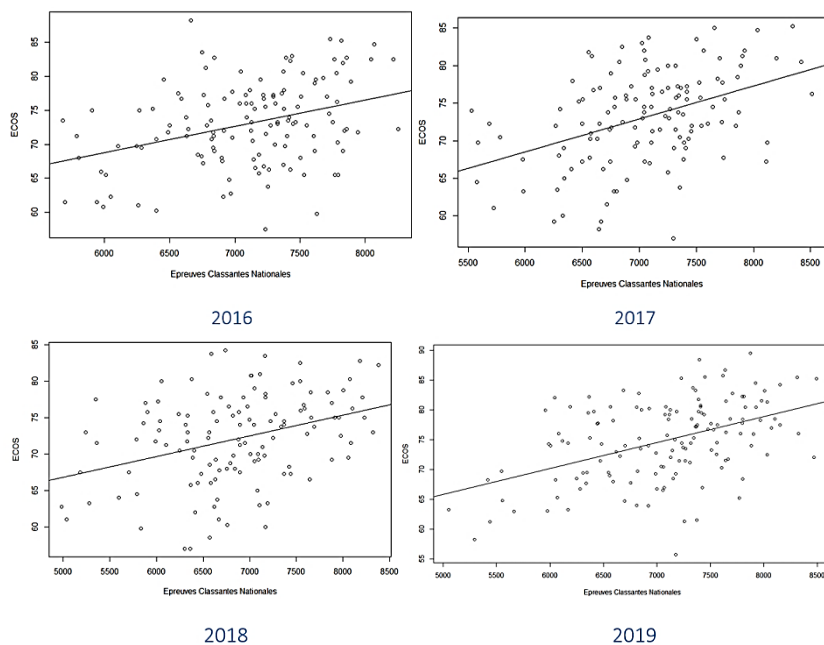
Il n'y a pas non plus de redondance avec les résultats aux épreuves théoriques. En effet, comme le montre la figure 5, le coefficient de corrélation entre les résultats à l'ECOS et les résultats de l'UE d'enseignements théoriques de 6<sup>e</sup> année est également significatif, positif et faible (0.40 [0.25;0.53] en 2016, 0.39 [0.23;0.53] en 2017, 0.25 [0.08;0.41] en 2018, 0.49 [0.36;0.61] en 2019) justifiant de l'intérêt de l'ECOS par rapport aux modalités d'évaluation classiques existantes. La modification des stations en 2018 apporte une tendance à la diminution de la corrélation mais ceci n'est pas vérifié sur les résultats 2019.



**Figure 5 • Nuages de points de l'ECOS et des épreuves théoriques de 6<sup>e</sup> année**

Enfin, il n'y a pas de redondance avec les ECN. En effet, le coefficient de corrélation entre les résultats à l'ECOS et les résultats des épreuves classantes nationales est lui aussi significatif, positif et faible (0.35 [0.19;0.49] en 2016, 0.43 [0.27;0.56] en 2017, 0.35 [0.19;0.50] en 2018, 0.46 [0.32;0.58] en 2019, cf. figure 6).

Ainsi l'examen ECOS présente une contribution spécifique qui peut être associée notamment à la mise en situation.



**Figure 6 • Nuage de points de l’ECOS et des épreuves classantes nationales**

Tous ces résultats mettent en évidence l’intérêt de la démarche en ce que l’évaluation qui en résulte n’est pas redondante par rapport à d’autres dispositifs d’évaluation. Nous allons maintenant prendre un peu de recul au regard de ce qui peut être appréhendé comme une évaluation priorisant des critères d’évaluation objectivables et standardisables dans une approche à dominante behavioriste.

### **3. Analyse critique**

#### **3.1. Compétences/critères d’évaluation : des catégorisations distinctes et complexes à mettre en cohérence**

L’annexe du BO (MESRI, 2013) décrit les compétences génériques requises en les articulant autour de 7 compétences (voir l’extrait du BO présenté en annexe de cet article). Mais il ne donne pas d’indication sur la façon dont ces compétences doivent être spécifiées et évaluées. C’est cette évaluation que vise cet examen, notamment concernant les compétences génériques de « clinicien » et « communicateur », tout au long des

7 stations. Quelles conceptions sont alors retenues des attendus et des modalités pour les atteindre ?

En termes de ressenti, 87 % des enseignants ont estimé que la formation actuelle permet aux étudiants d'acquérir les compétences requises. Cependant, il ne leur était pas demandé quelles étaient leurs définitions de ces compétences. Les étudiants sont beaucoup plus circonspects, quand on les interroge sur la maîtrise de chacune des 7 compétences génériques dans le cadre du questionnaire anonyme soumis après achèvement du parcours des stations d'évaluation : les taux de maîtrise déclarée varient de 47 % pour les compétences « d'acteur de santé publique » à 55 % pour celles de « coopérateur ». On retrouve cette tendance en 2018 : les taux de maîtrise déclarée varient de 50 % pour les compétences « d'acteur de santé publique » à 57 % pour celles de « coopérateur ». De même en 2019, les taux de maîtrise déclarée varient de 47 % pour les compétences « d'acteur de santé publique » à 64 % pour celles de « coopérateur ». Les étudiants concernés n'avaient découvert que tardivement ces différentes compétences génériques telles qu'énoncées, auxquelles ils n'avaient été sensibilisés que quelques semaines avant l'examen ; ces réponses traduisent également une difficulté de compréhension des attendus, tant les termes sont vastes et les définitions étendues.

La question se pose du passage des compétences à des actions susceptibles d'être évaluées. Selon Rey (2015), un référentiel (tel celui décrit dans le BO pour les compétences génériques) devient parole de commandement, mais ne garantit pas la faisabilité, surtout quand on aborde le caractère civique et social. Rey interroge la culture du résultat susceptible d'y être associée : il y a souvent, dans nombre d'approches par compétences, confusion entre la compétence et le résultat d'une action qui s'apparente alors à une performance. À titre d'exemple, les compétences génériques de communicateur requièrent de la part de l'étudiant qu'il soit à même « *d'établir une communication axée sur le patient, par le biais de la prise de décisions partagées et d'interactions efficaces fondées sur l'éthique et l'empathie* ». Or, il reste difficile de formuler des critères d'évaluation au-delà des formules de politesse et de civilité usuelles. Le BO laisse la mise en œuvre du référentiel sous la responsabilité des enseignants et des étudiants et fait quasi-silence sur le processus d'engendrement de ces compétences et de leur évaluation ; seules les modalités « d'évaluations en situation authentique et des auto-évaluations » sont évoquées.

### 3.2. Priorités aux critères formels et limites d'une approche behavioriste

Roegiers (2010) souligne l'importance d'établir une articulation entre le référentiel métier, le référentiel de compétences et le référentiel de formation initiale. Le Boterf (2011) ajoute la nécessité d'élaborer préalablement un référentiel d'évaluation. On sent bien que la volonté ministérielle d'introduire l'approche par compétences dans les études de médecine n'en est qu'au début de la démarche. La mise en œuvre de l'ECOS l'est également. Les critères d'évaluation ont été définis de manière empirique. La priorité donnée à l'opérationnel pour faire face aux défis de l'ingénierie pédagogique a conduit à mettre, en partie, de côté la réflexion sur les conceptions des relations soignants-patients et leurs modes d'enseignement.

Dans la définition de la compétence par Tardif, figure la notion de « *famille de situations* ». Cela interroge à deux niveaux par rapport à l'évaluation proposée : la répétition et la complexité. D'une part, cela sous-entend que la compétence ne peut être acquise que si elle a pu être mobilisée dans plusieurs situations. Quel crédit peut-on alors donner à une certification qui ne repose que sur une situation unique (celle de l'examen) ? Une réflexion est actuellement en cours pour étendre le dispositif d'évaluation (formative et sommative) aux situations de stage de la 2<sup>e</sup> à la 5<sup>e</sup> année de formation. D'autre part, dans une famille de situations, certaines seront plus complexes à traiter que d'autres. On se retrouve alors face à un paradoxe : les « vignettes » sont construites sur des situations simples et courantes du point de vue des médecins pour répondre aux attendus en fin de 2<sup>e</sup> cycle. Mais cela ne conduit-il pas à une conception réductrice de ce que seraient les compétences cliniques, recentrées sur l'interrogatoire, l'examen et la prescription sur un patient standardisé objet des soins ?

Il est en cela utile d'interroger la posture behavioriste qui sert souvent de fondement à l'approche par compétences. Comme le souligne Boutin, le behaviorisme est adapté pour la maîtrise de comportements observables, objectivés. Dans cette approche, « *le comportement visé doit être répété jusqu'à ce qu'il devienne automatique* » (Boutin, 2000). Ainsi en est-il lors d'examen cliniques de l'identification de certains signes cliniques. Leung montre les dérives d'une approche qui peut, dans son principe de répétitivité étendue à l'évaluation, tendre à l'absurdité (Leung, 2002) : il donne l'exemple d'un étudiant qui doit montrer au moins 4 fois qu'une

compétence est mise en œuvre au cours de 7 situations filmées différentes. Meirieu (2005) critique ce qui peut se traduire par l'atomisation des apprentissages globaux en une multitude de petits actes dont la recomposition peut être difficile.

Hébrard (2011) souligne quant à lui le risque qu'il y aurait à privilégier l'apprentissage par l'action en situation, au détriment des connaissances, des concepts et cadres d'analyse relevant des disciplines. Dans la formation d'infirmier, il constate que les grilles d'évaluation des compétences sont souvent hétérogènes et sont constituées d'une liste de preuves à vérifier ; les évaluateurs font alors face à la difficulté de cocher telle ou telle case. De plus, les dimensions affectives, conflictuelles, la question du pouvoir et les enjeux identitaires sont occultés.

Dans les études médicales, les connaissances issues des sciences sociales sont en bonne part ignorées, l'enseignement étant essentiellement le fait de praticiens. Or, l'approche par compétences, si elle est ramenée à une liste d'activités, ne permet pas de prendre en compte ce qui est identifié comme une caractéristique clé du travail de thérapeute. Le travail de thérapeute est en effet caractérisé par une tension intrinsèque entre, d'une part, un mouvement nécessaire d'objectivation du patient pour interpréter les signes éventuels de la maladie, et d'autre part une capacité à « *se déprendre de sa position d'objectivation et se placer avec le patient dans une relation de vivant à vivant et de sujet à sujet* » (Lefève et Mino, 2011).

### **3.3. Spécificités de la relation thérapeutique et questions ouvertes sur l'évaluation**

S'agissant des activités adressées à autrui, Maubant et Piot (2011) relèvent combien elles supposent de pouvoir identifier la singularité des situations et la nécessité d'y agir avec une certaine autonomie. Dans ces activités, la dimension observable, objectivable, n'est qu'une manifestation partielle de toute une activité invisible, consistant à « *percevoir des informations pertinentes, les organiser et les interpréter pour leur donner du sens, prendre des décisions, vérifier l'adéquation des effets réels de l'activité avec les effets attendus et le cas échéant, mettre en place des processus de régulation* » (*Ibid.*, p.9). Les termes utilisés tels que savoir-être, comportement, attitude, capacité relationnelle, compétence sociale, etc. résistent à une définition en tant qu'objet d'apprentissage pour « *ce qui est généralement conçu comme des qualités humaines attribuées à des traits de personnalité souvent considérés implicitement comme innés* » (Hébrard, 2011), alors qu'il s'agit de qualités socialement construites, et donc culturellement

formées. Selon cet auteur, les « zones d'ombre » auxquelles se heurtent ces apprentissages tiennent « *au flou des cadres conceptuels sous-jacents, à l'influence d'un paradigme issu du behaviorisme et du pragmatisme* », et « *à l'insuffisance de la réflexion épistémologique et du questionnement éthique* » (*Ibid.*, p. 107).

Étudiant les référentiels de compétence et de formation, notamment les grilles d'évaluation, Hébrard constate que la complexité de la relation humaine tend à être « *réduite à sa surface communicationnelle et à l'utilisation de "techniques". La profondeur et l'ambiguïté, les enjeux identitaires, sociaux, politiques et éthiques semblent occultés ou du moins mis au second plan* » (*Ibid.*, p. 113). Les formateurs se trouvent pris dans la tension entre des supports très prescriptifs et détaillés, qui notamment rigidifient l'évaluation, et l'affirmation de l'autonomie et de la responsabilité des étudiants, en ce qu'ils doivent savoir appréhender chaque situation dans sa singularité et sa globalité.

Il apparaît que les orientations adoptées en termes de formation et d'évaluation ont des liens avec la conception retenue de l'activité thérapeutique : si les compétences cliniques reposent non seulement sur la connaissance des pathologies et des moyens d'investigation et de traitement, mais, de façon imbriquée, sur « *la compréhension fine de ce que peut signifier le fait d'être malade pour un individu et son entourage, ainsi que la connaissance des enjeux sociaux et institutionnels de cette expérience* » (Lefève et Mino, 2011), alors les situations de relation soignant-patient sont fondamentalement des situations qui réclament une réflexion sur et dans l'action, et une réflexion menée non sur le patient mais avec lui.

Dès lors il apparaît qu'une opération qui paraît aussi simple que de définir des listes de critères d'évaluation est en fait sous-tendue par une conception plus générique de ce qu'est, en l'occurrence ici, un thérapeute. Une approche par compétences ramenée à une liste d'attitudes peut s'avérer converger avec une approche objectivante des patients et instrumentale des relations, précisément parce qu'elle n'interroge pas les dispositions requises pour être à même de mener ce double mouvement d'objectivation et de construction d'un échange effectif avec un patient.

La démarche des ECOS est également répandue dans les pays anglo-saxons avec ses checklists et ses patients-standardisés. Elle apporte certes l'individualisation de l'apprentissage mais revient selon certains auteurs à démontrer la performance de l'étudiant, et ce, potentiellement au détriment du processus sous-jacent (Leung, 2002). L'approche peut être

perçue comme superficielle et parfois source de démotivation. Elle encourage les étudiants à faire ce qu'on attend d'eux plutôt que de prendre le risque de développer une distance critique. À cela s'ajoute le risque pour les examinateurs de définir des règles minimalistes de validation. L'autre risque souligné par Leung, la réduction du contenu pédagogique, est particulièrement sensible dans le contexte d'études médicales où les étudiants désertent les cours magistraux et où le primat donné aux QCM fait que toute évaluation doit se conformer aux modalités d'un QCM, quand bien même les connaissances en cause sont complexes et relèvent d'un paradigme différent de celui de l'*evidence-based medicine*, ce qui est le cas des connaissances issues des sciences sociales.

Il résulte de ces éléments qu'il apparaît particulièrement important de mener une réflexion approfondie, en valorisant la réflexivité de la démarche d'innovation pédagogique retenue et de l'environnement informatique qui lui sert de socle. Le risque a contrario serait de plaquer des conceptions instrumentales des compétences et de leur évaluation aux activités et relations thérapeutiques. Des pistes d'évolution intéressantes pourraient être trouvées dans une approche plus transversale, intégrant notamment les mises en situation en stage et dans des évolutions de l'EIAH incluant une participation plus active des étudiants.

#### **4. Conclusion et perspectives**

Cette communication rend compte des premiers résultats d'une démarche innovante en termes de formation aux compétences cliniques de futurs médecins et de leur évaluation. L'environnement informatique constitué permet d'une part d'équiper l'ingénierie pédagogique, et d'autre part, distinctement, de recueillir de façon anonyme les retours tant des enseignants que des étudiants.

L'étude des résultats de cet examen de compétences cliniques montre qu'il n'est pas redondant au regard des examens évaluant les compétences plus « théoriques » sur les maladies, les techniques diagnostiques et thérapeutiques et les sciences dites fondamentales associées. En cela, cet examen présente une contribution spécifique, dont la pleine signification reste à expliciter.

L'analyse critique proposée en complément permet d'ouvrir des pistes sur les questions que soulève cet examen. L'observation du dispositif en train de se faire permet d'interroger les paradigmes en présence. Il apparaît que, compte tenu des impératifs de calendrier et de conceptions implicites



sur les relations soignants-patients, l'approche behavioriste ait été dominante dans les premiers temps de ce dispositif.

En perspectives, la question est celle de dispositifs complémentaires ou renouvelés de formation et d'évaluation qui fassent place à une approche constructiviste susceptible d'aider à affronter des situations qui réclament une réflexion sur et dans l'action (Boutin, 2000). Dans cette perspective, la refonte des stages et de leur évaluation pourrait être l'occasion d'une réflexion sur la mise en place de démarches pédagogiques et d'évaluation accordant plus de place à cette réflexion sur et dans l'action, non seulement sur, mais aussi avec les patients. Il s'agit également de réfléchir à un dispositif permettant de mesurer la progression des compétences acquises au fil du cursus, voire au-delà si une articulation peut être établie avec la formation du troisième cycle voire la formation continue des praticiens.

Dans ce cadre, l'EIAH doit permettre de mesurer la progression des compétences acquises en cours de formation et de contribuer à la réingénierie pédagogique, tout en étant lui-même objet d'enquête et de réflexion dans la perspective interdisciplinaire retenue.

## **REMERCIEMENTS**

Cette recherche est financée grâce au soutien de l'Institut de Recherche en Santé Publique, IReSP-17-HSR-16. Elle a bénéficié de l'aide des partenaires financeurs de l'IReSP dans le cadre de l'appel à projets général 2017 Volet Services de Santé.

J.C. Basson (CRESCO EA 7419\*), E. Coeurdevey (CERTOP UMR 5044\*), E. Fournales (CERTOP UMR 5044\*), N. Haschar-Noé (CRESCO EA 7419\*), P. Marrast (CERTOP UMR 5044\*), T. Lang (INSERM UMR 1027\*\*), M. Kelly-Irving (INSERM UMR 1027\*), S. Lamy (INSERM UMR 1027\*), L. Bonneville (GRICO\*\*) et S. Grosjean (GRICO\*\*), participent également à cette recherche dans le cadre d'une collaboration menée au sein de l'IFERISS, Institut Fédératif d'Études et de Recherches Interdisciplinaires Santé Société.

\* Université Toulouse 3

\*\* Université d'Ottawa

## **RÉFÉRENCES**

Boutin, G. (2004). L'approche par compétences en éducation : un amalgame paradigmatique. *Connexions*, 1(81), 25-41.

Boutin, G. (2000). Le béhaviorisme et le constructivisme ou la guerre des paradigmes. *Québec français*, 119, 37-40.

Dessus, P. (2019). *Tutoriel - Analyser les items d'un questionnaire*. Université Grenoble Alpes. <http://espe-rtd-reflexpro.u-ga.fr/docs/scied-cours-qcm/fr/latest/tuto-analyse-items.html>.

Djouad, T. (2013). *Ingénierie des indicateurs d'activités à partir de traces modélisées pour un environnement informatique d'apprentissage humain* [thèse de doctorat]. Université Claude Bernard Lyon 1, Lyon, France.

Hébrard, P. (2011). L'humanité comme compétence ? Une zone d'ombre dans la professionnalisation aux métiers de l'interaction avec autrui. *Les sciences de l'éducation - Pour l'Ère nouvelle*, 2(44), 103-121.

Husson, F., Lê, S. et Pagès, J. (2016). *Analyse de données avec R*. Presses Universitaires de Rennes.

Khanz, K., Ramachandran, S., Gaunt, K. et Pushkar, P. (2013). The objective Structured Clinical Examination (OSCE): AMEE guide n° 81. Part I: An historical and theoretical perspective. *Medical Teacher*, 9, e1437-e1446.

Khanz, K., Ramachandran, S., Gaunt, K. et Pushkar, P. (2013). The objective Structured Clinical Examination (OSCE): AMEE guide n° 81. Part II: Organisation and administration. *Medical Teacher*, 9, e1447-1463.

Le Boterf, G. (2011). *Ingénierie et évaluation des compétences*. Eyrolles.

Lecoutre, J.-P. (2016). *Statistique et probabilités*. Dunod.

Lefève, C. et Mino, J.-C. (2011). Former de vrais thérapeutes : La place des sciences humaines et sociales dans les études de médecine. *Études*, 4142, 187-198.

Leung, W. C. (2002). Competency based medical training: review. *British Medical Journal*, 325(7366), 693-696.

Maubant, P. et Piot, T. (2011). Étude des processus de professionnalisation dans les métiers adressés à autrui. *Les sciences de l'éducation - Pour l'ère nouvelle*, 2(44), 7-11.

Meirieu, P. (2005). *Si la compétence n'existait pas, il faudrait l'inventer...* Site de l'auteur. Récupéré le 10 décembre 2020 de <https://www.meirieu.com/ARTICLES/competence-eps.pdf>

Ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation (MESRI) (2013). *Études médicales : Régime des études en vue du premier et du deuxième cycle*. Bulletin officiel n° 20 du 16 mai. [https://www.enseignementsup-recherche.gouv.fr/pid20536/bulletin-officiel.html?cid\\_bo=71544&cbo=1](https://www.enseignementsup-recherche.gouv.fr/pid20536/bulletin-officiel.html?cid_bo=71544&cbo=1)

Rey, B. (2015). La notion de compétence : usages et enjeux. *Le français d'aujourd'hui*, 4(191), 15-24.

Roegiers, X. (2010) *Des curricula pour la formation professionnelle initiale: la Pédagogie de l'intégration comme cadre de réflexion et d'action pour l'enseignement technique et professionnel*. De Boeck Université.

Tardif, J. (2006). *L'évaluation des compétences: Documenter le parcours de développement*. Chenelière Éducation.

## ANNEXE

Extrait du BO (MESRI) n°20 du 16 mai 2013. « Études Médicales, Régime des études en vue du premier et du deuxième cycle ».

[https://www.enseignementsup-recherche.gouv.fr/pid20536/bulletin-officiel.html?cid\\_bo=71544&cbo=1](https://www.enseignementsup-recherche.gouv.fr/pid20536/bulletin-officiel.html?cid_bo=71544&cbo=1)

### Les compétences génériques

Les compétences dont l'acquisition est nécessaire sont celles de :

#### A) Clinicien

L'étudiant utilise son savoir médical et ses habiletés pour analyser, à partir du contexte de soins et de la plainte du patient, une situation clinique et pour dispenser des soins dans une approche centrée sur les patients.

Il :

- démontre sa capacité d'explorer les antécédents du patient et de les consigner de façon exacte, concise et organisée dans divers contextes cliniques ;
- réalise l'examen somatique adapté à la situation clinique et au projet de soins de patients de tout âge, et les principaux gestes techniques associés ;
- démontre sa capacité à mener un raisonnement hypothético-déductif en mobilisant ses connaissances des processus anatomo-physio-pathologiques ;
- développe les procédures diagnostiques des pathologies et urgences les plus fréquentes et sait amorcer le traitement approprié ;
- tient compte dans sa démarche décisionnelle des souhaits, des représentations et des compréhensions du patient et de son environnement éthique et culturel ;
- repère, dans une optique de prévention, les personnes exposées à des problèmes de santé courants qui mettent leur santé ou leur vie en danger ;
- développe une capacité d'observation globale de la situation ;
- est en mesure de présenter oralement de manière claire et concise sa démarche clinique auprès de l'ensemble de ses interlocuteurs.

#### B) Communicateur

Échangeant de façon dynamique avec le patient et son entourage, collaborant avec les différents professionnels du système de santé, l'étudiant a conscience des enjeux de la relation et de la communication verbale et non verbale sur la qualité des soins.

Il :

- établit une communication axée sur le patient, par le biais de la prise de décisions partagées et d'interactions efficaces fondées sur l'éthique et l'empathie ;
- obtient les renseignements pertinents et les points de vue du patient, de son entourage, des collègues et des soignants, et sait en faire la synthèse ;
- apprend à gérer son stress et celui des autres acteurs ;
- discerne quelle information doit être délivrée au patient, à son entourage, aux collègues et aux soignants, de manière claire, loyale et appropriée à leur niveau de compréhension et à leur culture, en s'appuyant sur les recommandations de bonne pratique ;
- s'appuie sur les nouvelles technologies de l'information.

#### C) Coopérateur, membre d'une équipe soignante pluriprofessionnelle

L'étudiant travaille en partenariat avec d'autres intervenants qui participent de manière appropriée au soin du patient.

(...)

**D) Acteur de santé publique**

L'étudiant participe à la vie des structures de soins et à la bonne gestion des ressources. Il contribue à l'efficacité et à l'efficience du système de soins pour améliorer l'état de santé global des patients et des populations.

(...)

**E) Scientifique**

L'étudiant comprend que pendant toute sa vie professionnelle il devra remettre en question et chercher à réactualiser ses connaissances afin de garantir son domaine d'expertise. Il comprend l'intérêt de la démarche scientifique pour élaborer de nouveaux savoirs.

(...)

**F) Responsable aux plans éthique et déontologique**

L'étudiant a une attitude guidée par l'éthique, le code de déontologie et adopte un comportement responsable, approprié, intègre, altruiste visant au bien-être personnel et à la promotion du bien public se préparant ainsi au professionnalisme.

(...)

L'étudiant doit également apprendre à être :

**G) Réflexif**

L'étudiant doit développer tout au long de son parcours d'apprentissage sa capacité d'auto évaluation et celle de se poser des questions pertinentes en situation réelle de soins et de prévention, en tenant compte du contexte clinique, institutionnel, légal et sociétal. Il doit démontrer sa capacité à se remettre en question et à argumenter ses décisions.

(...)



## Fonder un jeu sérieux sur un jouet-micromonde

► **Bertrand MARNE** (ICAR UMR 5191, Université Lumière Lyon 2, France)

---

---

■ **RÉSUMÉ** • Ryan, Costello et Stapleton proposent de fonder les jeux sérieux d'apprentissage sur un jouet systémique comparable à un micromonde. Selon leurs travaux, c'est le moyen pour concevoir des jeux sérieux efficaces, permettant un apprentissage ludique par l'exploration et l'expérimentation (Ryan *et al.*, 2012). Nous présentons dans cet article une mise à l'épreuve en situation écologique (23 prototypes de jeux conçus) de leur méthode de conception. Nous avons identifié plusieurs verrous à l'utilisation de cette méthode et nous proposons des solutions pour les lever, notamment un patron de conception de jeux sérieux fondés sur des jouets-micromondes.

■ **MOTS-CLÉS** • Apprentissage par le jeu, jeu sérieux, apprentissage par l'investigation, micromondes, jeux de simulation, patron de conception.

■ **ABSTRACT** • Ryan, Costello et Stapleton invite basing design of a serious learning games on a systemic toy, close to the concept of microworld. Their work points out that it is the best way to build entertaining serious games allowing exploration and experimentation (Ryan *et al.*, 2012). In this paper we report an ecological evaluation (23 games prototype made) of their design method. We identify several issues and propose some solutions, including a full design pattern to build serious games based on toy-microworld.

■ **KEYWORDS** • Game-based learning, serious games, inquiry learning, microworlds, simulation games, design patterns.

## **1. Introduction**

Nous choisissons de focaliser cette étude sur les Jeux Sérieux (JS) d'apprentissage cherchant à obtenir une motivation intrinsèque des apprenants-joueurs plutôt que sur les JS ayant recours à la motivation extrinsèque par le biais de mécanismes de *gamification*. Nous visons plus précisément les JS qui ne cherchent pas seulement à augmenter la motivation, mais ceux dans lesquels les interactions sont utilisées pour proposer aux apprenants-joueurs de l'exploration et de l'expérimentation et qui suscitent un apprentissage plus constructiviste que transmissif. Ryan, Costello et Stapleton, en s'appuyant explicitement sur les travaux de Schell (2008) et dans une réinterprétation implicite du concept de micromonde (Papert, 1987), proposent de centrer la conception de cette catégorie de JS sur la construction d'un *jouet systémique* (Ryan *et al.*, 2012). Nos travaux s'inscrivent dans cette problématique générale qui interroge les apports de la notion de jouet dans les méthodes de conception des JS.

Dans cet article, nous décrivons nos travaux, constitués de trois études de terrain qui évaluent les méthodes de conception de Ryan *et al.*, et qui permettent d'extraire deux patrons de conception (Alexander *et al.*, 1977 ; Meszaros et Doble, 1997) de JS articulés sur des jouets systémiques.

Nous présentons d'abord, en section 2, les travaux antérieurs qui nous ont servi de base pour définir les concepts de notre problématique, et notamment comment cette notion de jouet systémique rejoint celle des micromondes (Papert, 1987). Dans la section 3, nous exposons la méthodologie de recherche que nous avons choisie et les biais qu'elle provoque pour mener les trois études (itérations) que nous décrivons et discutons en section 4. Nous présentons nos principaux résultats en section 5, avant de conclure.

## **2. Le jouet comme fondement de jeux sérieux constructivistes**

La mise en avant des JS comme outils pédagogiques est souvent justifiée par la motivation et l'engagement qu'ils sont susceptibles de provoquer (Garris *et al.*, 2002 ; Keller, 1995 ; Malone et Lepper, 1987 ; Prensky, 2004). Pourtant, selon certains travaux, cette utilisation des JS pour motiver a aussi été critiquée (Wouters *et al.*, 2013). Selon cette méta-étude analysant 39 travaux de recherche, les JS ne sont pas sensiblement plus motivants que les autres outils d'enseignement utilisés dans les groupes témoins, et ce

en tenant compte du fait que les auteurs reconnaissent la difficulté de mesurer correctement la motivation.

Dans le même ordre d'idée, certains auteurs comme Bogost et Ryan *et al.* considèrent que l'intérêt principal des JS pour l'apprentissage n'est pas la motivation, mais de proposer un apprentissage expérientiel (Bogost, 2007 ; Bogost, 2013 ; Ryan *et al.*, 2012). En effet, ils montrent que les JS de ce type proposent des expériences d'apprentissage plus constructivistes, qui à leur tour peuvent induire une motivation intrinsèque. Alors que d'autres JS, articulés sur les principes de gamification, tentent parfois de n'apporter qu'une motivation extrinsèque (Bogost, 2013). Or, dans certains de ces cas, la gamification peut aussi avoir des effets négatifs sur l'apprentissage, comme la baisse de la motivation intrinsèque (Deci *et al.*, 1999) ou la course à la récompense (Callan *et al.*, 2015).

Les travaux de Ryan *et al.* (2012) en utilisant la « *Procedural Rethoric* » (rhétorique de la procédure)<sup>1</sup> proposée par Bogost (2007) et le « *Deep Conceptual Learning* » (apprentissage conceptuel profond)<sup>2</sup> proposé par Gee (2009) invitent à fonder la conception d'un JS sur l'utilisation d'un jouet (numérique) grâce à la « *Lens of the Toy* » (objectif du jouet) de Schell (2008). Selon Ryan *et al.*, un bon jouet est un système complexe qui propose de nombreuses affordances ludiques (Brougère, 2013) et qui sollicite les capacités cognitives de reconnaissance de structures, de raisonnement stratégique et de résolution de problèmes. En cela, ces auteurs enrichissent la notion de jouet proposée par Schell. Ce dernier définit moins le jouet par la description de l'artefact lui-même, que par l'activité qu'il engendre : « le jouet est artefact avec lequel on joue », et non l'artefact auquel on joue, qui est lui plutôt le jeu ; l'objet-jeu se décrit alors comme l'ensemble de règles (obstacles et buts) qui encadrent l'utilisation du jouet (Schell, 2008). De plus, Schell considère déjà qu'un jeu conçu autour d'un jouet est une activité de résolution de problèmes envisagée avec une attitude ludique.

Ainsi, Ryan *et al.*, sans l'explicitier dans leurs travaux, retrouvent des notions proches des micromondes, puisqu'ils considèrent qu'un jouet, au cœur d'un JS éducatif, doit être un modèle concret des systèmes (une simulation) qui dirigent le domaine à enseigner (Papert, 1987 ; Ryan *et al.*, 2012). Et ainsi, ils décrivent le jouet comme un outil qui invite, par l'action

---

1 Capacité d'un processus interactif à porter un message.

2 Un apprentissage profond est émergent et se fait intrinsèquement par l'activité de jeu.

de jeu, à l'exploration des systèmes, à la reconnaissance de ses structures et à l'expérimentation pour essayer de contrôler, puis de maîtriser ces systèmes. C'est par ces moyens que Ryan *et al.* considèrent qu'on atteint le « *Deep Conceptual Learning* » décrit par Gee (2009).

Ces conceptions sont très proches de celles qui fondent le concept de micromonde. Les micromondes sont des EIAH qui permettent de mettre les utilisateurs dans une situation d'autonomie afin de développer des connaissances par la découverte, la conception d'hypothèses et l'expérimentation de celles-ci (Papert, 1987). Ainsi, lorsqu'un apprenant-joueur manipule un micromonde ou un jouet selon Ryan *et al.*, il cherche à comprendre les règles et les équilibres existant dans les systèmes que le jouet présente, et ce de deux manières. Soit en l'explorant : il provoque des causes et en observe les conséquences pour en induire des règles. Soit en expérimentant : il cherche à obtenir des conséquences en multipliant les possibilités de causes pour vérifier une ou plusieurs règles, ses hypothèses, par abduction (Peirce, 1902).

Dans leur interprétation de ces concepts, Ryan *et al.* (2012) identifient trois états cognitifs que le jouet doit créer chez l'apprenant-joueur. L'« *abstraction* » provoquée par la confrontation avec un système concret implémente des concepts abstraits et s'apparente à une simulation du domaine à enseigner. Cet état permet l'exploration et autorise à l'apprenant-joueur la construction de modèles personnels de représentation, en étudiant les structures des systèmes auxquels il est confronté et en raisonnant par induction et abduction. L'état de « *control* » est obtenu, grâce à des interfaces ludiques, avec la simulation du domaine. L'idée est de parvenir à l'aphorisme attribué à l'un des pionniers du jeu vidéo Nolan Bushnell : « facile à prendre en main, difficile à maîtriser » (Crawford, 1984), tout en maintenant le plaisir de l'utilisation. Le troisième état est l'appropriation, l'« *ownership* », dans lequel l'apprenant-joueur, ayant pu expérimenter ses modèles hypothétiques, renforce ses compétences en travaillant à leur maîtrise.

Pour construire ces JS fondés sur des jouets/micromondes, Ryan *et al.* proposent, sans la vérifier expérimentalement, une méthode qui ressemble beaucoup à certains de nos travaux (légèrement antérieurs) : les 6 facettes de la conception des jeux sérieux (Marne, 2014 ; Marne *et al.*, 2011). Les 6 facettes forment un modèle formel de la conception des JS qui a été conçu pour faciliter la collaboration entre les parties prenantes de cette conception (Marne, 2014) et qui a été évalué expérimentalement.



La méthode de Ryan *et al.* se décompose en 5 étapes chronologiques (contrairement aux 6 facettes qui ne sont pas chronologiques).

1. *Identifier les mécaniques et dynamiques impliquées dans les systèmes du domaine que l'on cherche à enseigner.* Cette étape correspond de façon remarquable à deux facettes : la facette 1 (*Définition des Objectifs pédagogiques*), qui correspond à la modélisation du domaine à enseigner décrit à partir des objectifs des concepteurs, et la facette 2 (*Simulation du domaine*), qui décrit les modèles formels des systèmes à l'œuvre dans le domaine à enseigner.

2. *Présenter le système de façon à faciliter la reconnaissance de ses structures sous-jacentes.* Ceci rejoint les concepts décrits dans la facette 2 (*Simulation du domaine*). Ryan *et al.* proposent de s'inspirer des « Machinations » de Dormans, qui sont des patrons de conception pour les mécaniques de jeux vidéo (Dormans, 2012).

3. *Fournir un outil pour que l'apprenant-joueur ait un contrôle incarné et ludique.* Il s'agit donc d'un outil à la fois simple et aux affordances multiples permettant une interaction fine et ludique avec le système. Il s'agit exactement de ce que propose la facette 3 (*Interactions avec la simulation*), qui concerne la conception de l'interface ludique avec la simulation (facette 2) proposée à l'apprenant-joueur.

4. *Ajouter des buts pour créer des étapes dans la découverte du système par les apprenants-joueurs.* C'est en correspondance avec la facette 4 (*Problèmes et progression*), qui concerne les défis et les obstacles à proposer aux apprenants-joueurs afin de les faire progresser à la fois dans le jeu et dans les compétences recherchées.

5. *Fournir le nécessaire pour permettre aux apprenants-joueurs de partager leurs nouvelles compétences.* Ceci ressemble à la facette 6 (*Conditions d'utilisation*), qui concerne notamment l'élaboration des conditions d'institutionnalisation des apprentissages.

Assez logiquement, compte tenu de leurs propositions en opposition à certaines techniques de gamification, les étapes proposées par Ryan *et al.* ne reprennent pas les contenus de la facette 5 (*Décorum*) correspondant à la conception des éléments de gamification d'un JS.

En reprenant la définition du jouet donnée dans la « *Lens of the Toy* » de Schell (2008), nous avons constaté que parmi ces cinq étapes, seules les trois premières sont centrées sur l'élaboration du jouet. Les deux suivantes sont centrées sur la construction de l'objet-jeu.

En nous fondant sur les travaux de Ryan *et al.*, nous décrivons le « jouet systémique », ce jouet/micromonde qui sert de base à un jeu sérieux, comme ayant 4 caractéristiques :

- il permet de jouer (manipuler avec plaisir) sans but ;
- il permet d'explorer ;
- il permet d'expérimenter ;
- et surtout, il permet que cette manipulation, cette exploration et cette expérimentation soient pratiquées sur le domaine « sérieux » visé par le jeu.

La problématique générale de nos travaux est de questionner les JS fondés sur des jouets systémiques. Nous explorons notamment l'hypothèse que ces JS, comme les micromondes (Papert, 1987), permettent aux apprenants-joueurs, d'une part, la découverte par l'exploration, la compréhension et la modélisation des systèmes qui régissent les domaines enseignés par des raisonnements inductifs et, d'autre part, l'expérimentation pour vérifier et conforter les modèles hypothétiques par des raisonnements abductifs.

Dans cette perspective, les travaux présentés ici interrogent la conception des JS de cette catégorie. Ainsi, notre contribution est d'avoir mis à l'épreuve sur le terrain les propositions méthodologiques de Ryan *et al.*, et d'avoir extrait des éléments de deux patrons de conception (Alexander *et al.*, 1977 ; Meszaros et Doble, 1997) pour ce type de JS.

### **3. Propositions méthodologiques**

Pour proposer et mettre à l'épreuve des méthodes de conception de JS fondés sur des jouets, nous avons choisi de nous inscrire dans une démarche de recherche écologique dans la lignée de la recherche-action (Hugon et Seibel, 1988). Ce choix est guidé à la fois par des choix épistémologiques et par des contraintes pratiques.

S'agissant des choix épistémologiques, nous cherchons à extraire des éléments de patrons de conception pour aider à la fabrication des JS centrés sur des jouets. Dans le domaine des JS, les méthodes récentes d'extraction de patrons de conception s'intéressent à leur extraction à partir de JS existants (Marfisi-Schottman et Piau-Toffolon, 2015). Ces méthodes

permettent d'identifier les bons résultats et d'en inférer des patrons de conception, mais ne permettent malheureusement pas d'identifier les obstacles et problèmes survenus lors de la conception, et les moyens de les résoudre. C'est pourquoi, en complément, nous avons choisi de travailler sur une extraction de patrons *durant* les itérations du processus de conception lui-même. Ceci implique une approche très écologique, comme celles des méthodes de la recherche-action, en suivant les concepteurs aussi souvent que possible.

De plus, comme les JS fondés sur les jouets sont rares, nous avons aussi choisi d'être partie prenante, pour aider les concepteurs que nous suivions en construisant avec eux, quand c'était nécessaire, des éléments méthodologiques notamment élaborés à partir des expériences antérieures et de l'état de l'art. Ainsi, nous avons choisi de nous inscrire dans une épistémologie proche de la recherche Collaborative Orientée par la Conception (ROC) (Sanchez et Monod-Ansaldi, 2015), elle-même inspirée de la *Design Based Research* (Wang et Hannafin, 2005).

Ce choix nous a également permis de surmonter l'une des principales contraintes pratiques de ce type de recherche : trouver un terrain d'étude dans lequel la recherche peut être suffisamment longue pour permettre d'observer et d'agir sur des situations variées, sans que le risque économique pesant sur les concepteurs soit grand. Dans ce contexte proche de la ROC, nous avons choisi le cadre pédagogique d'un enseignement universitaire sur les JS.

Bien que ces choix méthodologiques aient rendu possible une étude d'une certaine ampleur par sa durée et son nombre de participants, nous avons dû tenir compte d'un certain nombre de biais : par exemple, le fait que nous soyons partie prenante, ou le fait que ce public qui découvre les JS est peut-être assez différent du public que nous pourrions cibler avec des patrons de conception.

Pour prendre en compte ces biais, au-delà de l'analyse des JS produits (Marfisi-Schottman et Piau-Toffolon, 2015), de celle des prototypes conçus à chaque itération de la production, et des observations directes des méthodes employées, nous avons aussi collecté des réponses à des questionnaires et à des entretiens proposés à trois occasions : avant le début de la conception et des discussions méthodologiques, pendant la conception, et après la production du JS final et son évaluation scolaire. Lors de toutes ces observations, notre collecte d'informations a été guidée par plusieurs objectifs. Nous avons cherché à savoir quelles étaient les

préconceptions des étudiants sur les méthodes de conception de JS et comment et pourquoi ces méthodes ont pu changer au cours de leur projet, afin d'avoir, pour chacun des participants, une idée des connaissances antérieures, des choix et des sources d'information qui ont participé à l'évolution de ces méthodes de conception au fil du temps.

Au cours de nos trois études, nous n'avons pas cherché à évaluer l'efficacité pédagogique ou didactique des JS produits. Notre objectif était de savoir si la méthode de conception de Ryan (centrée sur celle du jouet systémique du JS) permet de créer des JS induisant l'exploration et l'expérimentation, et d'en extraire des patrons de conception. Par conséquent, nous avons surtout cherché à savoir si, et comment, les JS produits avaient ces caractéristiques. Pour cela, nous nous sommes notamment fondés sur les qualités du JS (présence d'une simulation, d'une interface ludique) et sur les résultats des séances de *playtests*<sup>3</sup> (Davis *et al.*, 2005 ; Drachen *et al.*, 2018) menées par les participants sur des apprenants-joueurs extérieurs à nos trois études.

Nous avons travaillé en trois étapes (itérations), avec plusieurs groupes d'étudiants. Progressivement, nous avons amélioré les aides méthodologiques fournies en nous appuyant sur les difficultés des étudiants à appréhender et à mettre en œuvre la notion de jouet.

#### **4. Étude de terrain et discussion**

Cette étude de terrain s'est déroulée sur une période de 3 ans (une itération par an) auprès de 104 étudiants, qui ont développé 23 prototypes de JS, par groupes de 3 à 12 personnes. 58 de ces étudiants étaient en Master 2 (M2) de journalisme dans un cursus axé sur les pratiques numériques journalistiques. Les 46 autres étudiants étaient en Diplôme Universitaire (DU) de conception de jeu (level design), niveau bac+3.

Les étudiants ont travaillé à la conception et à la réalisation de prototypes dans le cadre de leur formation aux JS en général, et aux *newsgames* (Bogost *et al.*, 2012) en particulier pour les étudiants journalistes. Ces projets sont menés de façon agile, durant un semestre, avec des cours méthodologiques (entre 12 et 24 heures) auxquels s'ajoute de l'accompagnement.

---

<sup>3</sup> Technique d'évaluation des jeux très employée dans l'industrie des jeux vidéo.

Les prototypes de JS issus de ces travaux restent pour la plupart assez rudimentaires, dans leurs aspects aussi bien pédagogiques que ludiques, voire dans le mélange de ces deux aspects. L'objectif n'était pas de faire concevoir un JS idéal, mais en produisant un premier prototype de JS, de permettre aux étudiants de se familiariser avec ce type de média. Ainsi, comme nous l'évoquons dans la section précédente, nous ne décrivons ni ne discutons des qualités ludiques ou pédagogiques des projets portés par ces étudiants. Nous nous limitons à décrire et discuter la présence, la conception et l'intégration de jouets systémiques dans les JS produits.

Pour travailler, les étudiants bénéficiaient d'un bagage méthodologique et logiciel fourni lors des premiers cours. Ces contenus ont varié entre les itérations, en fonction des résultats précédents. Les trois études sont présentées dans les sous-sections suivantes.

#### **4.1. Première étude : le jouet présenté comme une des méthodes de conception**

La première étude menée a réuni 26 étudiants de M2 de journalisme, regroupés dans 6 projets de JS (dont un groupe de 9, pour un projet de plus grande ampleur) dont ils ont choisi le sujet librement.

##### **4.1.1. Protocole suivi**

Les étudiants ont bénéficié d'une introduction aux JS fondée sur *Journalism at Play* (Bogost *et al.*, 2012) et mettant en avant les *newsgames* comme des médias systémiques et expérientiels. Puis, la méthodologie des 6 facettes de la conception des jeux sérieux et leurs patrons de conception (Marne *et al.*, 2011) leur a été expliquée parce qu'elle décompose la conception de façon intelligible, en particulier pour des novices. Au cours de ces apports méthodologiques, la notion de jouet a été définie et présentée comme une bonne pratique pour combiner les facettes de la simulation du domaine et de l'interface ludique avec cette simulation.

Pour accompagner la réalisation technique, des outils auteurs ont été proposés avec une grande liberté de choix. Nous les classons en plusieurs catégories : un outil de conception, *Legadee*, cité par Marfisi-Schottman (2012), des outils de fabrication de narrations interactives (*Twinery* et *Klynt*), un outil de réalisation de jeux de rôles (RPG Maker), un outil auteur de JS *point & clic*, *eAdventure*, cité par Moreno-Ger *et al.* (2005), des outils de réalisation de jeux sans programmation textuelle (*GDevelop* et *Construct*).

### 4.1.2. Productions des étudiants

Les 6 travaux qui en résultent sont synthétisés dans le Tableau 1 et le Tableau 2. La colonne *Ambition* (cf. Tableau 1) décrit le but que les étudiants voulaient atteindre avec le JS. Dans le Tableau 2, la colonne *Genre de JS* choisi décrit à quel genre vidéoludique appartient le JS, la colonne *Principales mécaniques choisies* décrit les interactions et les moyens ludiques principaux mis en œuvre dans le JS. Enfin, la colonne *Méthode d'apprentissage visée* décrit comment les étudiants projetaient que se fasse l'institutionnalisation des savoirs dans leur JS.

**Tableau 1 • Descriptif général des JS produits lors de l'étude 1**

Nom du JS	Outil auteur utilisé	Ambition
<i>ISIS the End?</i> <sup>4</sup>	<i>Klynt</i>	Faire comprendre les mécanismes de la radicalisation
<i>Citizens of Nowhere</i>	<i>Twinery</i>	Découvrir les épreuves endurées par les réfugiés
<i>Dealopolis</i>	<i>RPG Maker</i>	Comprendre l'articulation drogue-corruption
<i>Femmes à travers le monde</i>	<i>Construct</i>	Sensibilisation aux risques d'agression pour les femmes
<i>Reportage Ebola</i>	<i>Twinery</i>	Comprendre les conditions du journalisme africain
<i>Vélo en ville</i>	<i>Twinery</i>	Sensibiliser aux dangers de la circulation pour les cyclistes

Leurs travaux n'ont pour la plupart pas dépassé le stade de prototypes et à ce titre n'ont pu être évalués que lors de *playtests* rudimentaires. Au cours de ceux-ci, l'institutionnalisation des savoirs liés à l'utilisation du JS n'a pas pu être mesurée.

### 4.1.3. Observations

Les observations rapportées ici pour chaque itération portent sur trois aspects : les caractéristiques du JS produit, les éléments remarquables de son processus de conception/développement, et les résultats des *playtests* menés. L'objectif de ces observations est de répondre à trois questions :

- Est-ce qu'un jouet systémique est intégré dans le JS ?
- Comment le processus de conception a, ou n'a pas, permis cette intégration ?

---

<sup>4</sup> ISIS the End? est un projet de plus grande ampleur (projet de l'année des étudiants) qui a été publié en ligne par la suite (James *et al.*, 2016).

– Et enfin, est-ce que cette intégration (quand elle a eu lieu) a permis de vérifier lors des *playtests* son utilisation comme un jouet systémique, c'est-à-dire avec ces trois caractéristiques : il permet de jouer (manipuler avec plaisir) sans but, il permet l'exploration, il permet l'expérimentation ?

**Tableau 2 • Synthèse des buts et des moyens mis en œuvre dans les JS produits lors de l'étude 1**

Nom du JS	Genre de JS choisi	Principales mécaniques choisies	Méthode d'apprentissage visée
<i>ISIS the End?</i>	Narration interactive	À chaque étape un choix détermine l'étape suivante	Exploration + texte de remédiation quand il y a erreur
<i>Citizens of Nowhere</i>	Narration interactive	À chaque étape un choix détermine l'étape suivante. Des choix font varier des jauges déclenchant des événements	Exploration + expérimentation des hypothèses
<i>Dealopolis</i>	Narration interactive	Collecte d'indices dans une carte. Puis QCM.	Exploration (indices, réponses au QCM) + textes de remédiation
<i>Femmes à travers le monde</i>	Associations d'objets	Association de choix en fonction de la situation	Le joueur perd toujours : Exploration + texte de remédiation
<i>Reportage Ebola</i>	Narration interactive	À chaque étape un choix détermine l'étape suivante	Exploration + textes de remédiation
<i>Vélo en ville</i>	Narration interactive	À chaque étape un choix détermine l'étape suivante	Exploration + textes de remédiation

Notre première observation est que les projets des étudiants sont quasiment tous des projets de narrations interactives, dans lesquelles les interactions se résument le plus souvent à faire des choix qui ne font que brancher la narration vers une succession linéaire et limitée d'autres étapes de narration. De plus, les choix proposés sont souvent binaires et plutôt manichéens que cornéliens : entre bon et mauvais choix (parfois évidents l'un comme l'autre).

## Bertrand MARNE

Avec ces supports, les étudiants avaient prévu que l'institutionnalisation des savoirs se ferait soit par exploration et remédiation : en testant un choix, sans connaissance préalable, qui s'il est « mauvais » mène à des documents de remédiation, et à la possibilité de refaire ce même choix, ou un choix très similaire. Soit par instruction et éventuellement remédiation : en lisant des « indices » fournis avant les choix et en vérifiant avec ce choix que l'indice est compris, ou en accédant à de la remédiation dans le cas contraire.

Un projet se distingue : *Citizens of Nowhere* parce qu'en plus de la narration interactive, il ajoute un système de jauges (santé, argent) qui varie en fonction des choix et qui peut réaiguillier la narration. Ainsi, la narration initialement arborescente a été rendue plus systémique. Tardivement dans la conception, les étudiants avaient voulu mettre en place une petite simulation au cœur de leur JS très narratif. Leur souhait était d'utiliser la notion de jouet systémique pour provoquer, en plus du couple exploration-remédiation, un apprentissage par l'expérimentation.

Les observations menées tout au long du processus de conception nous ont permis d'observer que cette décision s'est imposée assez progressivement chez ces étudiants. Notamment lorsqu'ils ont voulu mettre en place une jauge d'argent pour rendre compte de la pauvreté des migrants, puis dans un second temps une jauge de santé pour montrer leur précarité médicale. C'est ainsi que les étudiants se sont rendu compte lors de *playtests* que les apprenants-joueurs trouvaient intéressant et instructif de jongler avec les jauges en testant l'effet de différentes actions (qui avaient alors un effet bien plus rudimentaire que dans la version définitive).

Nous avons d'ailleurs retenu l'idée de *jauge* pour modéliser les variables à l'œuvre dans la simulation des systèmes, car d'après nos observations de phases de conception et nos entretiens finaux, il a semblé parlant pour l'ensemble des étudiants.

À l'exception de ce groupe auteur de *Citizens of Nowhere*, nous avons constaté que les étudiants n'avaient pas fait le pas de côté suggéré par Bogost *et al.* (2012). Ces derniers proposent d'éviter les méthodes journalistiques narratives pour fournir des contenus plus expérientiels et systémiques, dans lesquels l'apprentissage, émergent, se fait par des combinaisons d'exploration et d'expérimentation. Du point de vue méthodologique, nous n'avions donc pas réussi à proposer des moyens qui permettaient aux étudiants d'envisager le jouet comme solution valide



pour construire des JS avec une représentation systémique du domaine et des expériences avec ce type d'apprentissage, plutôt que des JS fondés sur les récits journalistiques auxquels ils ont été habitués par le reste de leur formation.

Nous pensons également que le choix des outils auteurs les plus simples à prendre en main (*Twinery*, *Klynt* et *RPG Maker*) par ces étudiants a joué un rôle important dans le choix des contenus puisqu'ils sont conçus pour les narrations interactives.

Nous avons utilisé ces observations pour construire et mener notre deuxième étude.

## **4.2. Deuxième étude : le jouet placé au centre des méthodes de conception**

Notre deuxième itération s'est déroulée d'une part auprès de 32 étudiants en M2 de journalisme (regroupés dans 8 projets de JS) et d'autre part auprès de 23 étudiants en DU de *level design* (regroupés dans 5 projets de JS).

### **4.2.1. Protocole pour les étudiants journalistes**

Pour les étudiants journalistes, la mise en œuvre a été fondée sur les mêmes principes que ceux de l'étude précédente, avec les changements suivants : le concept de jouet a été placé au centre des enseignements méthodologiques, sans pour autant en proposer une méthode de conception détaillée. Son articulation avec les *systèmes* décrits dans (Bogost *et al.*, 2012) est réifiée grâce au principe des *jauges*, retenu à la précédente itération de l'étude. Elles permettent de modéliser les variables systémiques dont le jouet peut faire varier les valeurs.

Afin d'éviter l'influence trop forte des outils auteurs dans le genre du JS, nous avons demandé aux étudiants journalistes, dans un premier temps, de faire un *playtest* avec un prototype non numérique (jeu physique, jeu de plateau, etc.) et, dans un second temps seulement, de concevoir un prototype numérique pour la fabrication duquel les outils auteurs les plus polyvalents (*GDevelop* et *Construct*) ont été proposés en priorité.

### **4.2.2. Productions des étudiants journalistes**

Les travaux de huit groupes sont présentés dans le Tableau 3 et le Tableau 4.

**Tableau 3 • Descriptif général des JS produits par les étudiants journalistes lors de l'étude 2**

<b>Nom du JS</b>	<b>Outil auteur utilisé</b>	<b>Ambition</b>
<i>Enfance dans le monde</i>	<i>GDevelop</i>	Sensibiliser aux difficultés des enfants des pays pauvres
<i>Jeu de l'Amazonie</i>	<i>GDevelop</i>	Comprendre les équilibres écologiques et économiques de l'agriculture en Amazonie
<i>La Catalogne</i>	<i>Construct</i>	Comprendre les enjeux de l'indépendance de la Catalogne
<i>Make Hummus not Walls</i>	<i>GDevelop</i>	Connaître des faits sur la Barrière de séparation d'Israël
<i>Make Twitter Great Again</i>	<i>Klynt</i>	Faire connaître les tweets du Président des USA
<i>Rentabilise Neymar</i>	<i>Twinery</i>	Comprendre les transferts et le management des joueurs de football
<i>Garde la pêche</i>	<i>Twinery</i>	Apprendre à avoir une alimentation équilibrée à faible impact écologique
<i>Un toit sur Lyon</i>	<i>GDevelop</i>	Apprendre les démarches pour se loger

**Tableau 4 • Synthèse des buts et des moyens mis en œuvre dans les JS produits par des étudiants journalistes lors de l'étude 2**

Nom du JS	Genre de JS choisi	Principales mécaniques choisies	Méthode d'apprentissage visée
<i>Enfance dans le monde</i>	Jeu d'adresse	Se déplacer en évitant de se faire attraper	Exploration + expérimentation
<i>Jeu de l'Amazonie</i>	Stratégie au tour par tour	Tour par tour : choix (déforestation, plantation, jachère, récolte, etc.) en fonction de ses ressources	Exploration + expérimentation
<i>La Catalogne</i>	Jeu d'action	Se déplacer avec agilité en collectant des informations	Lecture des informations
<i>Make Hummus not Walls</i>	Casse brique	Récupérer des messages par agilité	Lecture des informations
<i>Make Twitter Great Again</i>	Suite de QCM humoristiques	À chaque étape un QCM à remporter ou perdre	Exploration + textes de remédiation
<i>Rentabilise Neymar</i>	Narration interactive	À chaque étape un choix détermine l'étape suivante + système de jauges → ajouté à la fin	Exploration + texte de remédiation + expérimentation
<i>Garde la pêche</i>	Narration interactive	À chaque étape un choix détermine l'étape suivante + système de jauges → ajouté à la fin	Exploration + texte de remédiation + expérimentation
<i>Un toit sur Lyon</i>	Jeu de l'oie	Déplacement sur un plateau + questions à chaque étape	Exploration + textes de remédiation

Sur les productions des étudiants journalistes décrites dans le Tableau 3 et surtout dans le Tableau 4, nous observons que la proposition de JS a été plus diversifiée que lors de l'étude précédente.

Bien qu'il y ait encore deux JS qui sont des narrations interactives, il faut noter que, cette fois, les deux groupes concernés ont cherché à ajouter un système de jauges incrémentées/décrémentées en fonction des choix. Comme pour la première étude, il s'est agi d'un ajout tardif, en cours de conception, les étudiants cherchant absolument à intégrer un

jouet dans leur JS (c'était dans la consigne qui leur était donnée). Dans ces deux cas, à la manière des tests dans les magazines estivaux, le résultat obtenu dans ces jauges sert à classer l'apprenant-joueur dans une catégorie à la fin du JS.

D'autres JS conçus lors de cette étude n'ont pas d'aspect systémique. Le jeu de l'oie, le casse brique et le QCM humoristique ne proposent pas d'expérimentation aux apprenants-joueurs, seulement la lecture d'informations.

Dans *La Catalogne* et *Make Hummus not Walls* les étudiantes ont pourtant intégré un « jouet » au sens de Schell (2008), c'est-à-dire un élément amusant à utiliser, indépendamment des buts du JS lui-même. Mais, ce jouet était totalement décorrélié des aspects systémiques du domaine à découvrir. Il ne s'agissait pour ces groupes que de *gamification*, donc avec un but de motivation totalement indépendant des apprentissages. Par conséquent, il ne s'agit pas du tout d'un jouet au sens de Ryan *et al.*, c'est-à-dire utilisable comme un micromonde pour l'apprentissage par l'exploration et l'expérimentation (Papert, 1987 ; Ryan *et al.*, 2012).

Seuls les deux JS *Enfance dans le monde* et *Jeu de l'Amazonie* proposent aux joueurs de comprendre comment fonctionnent des systèmes en leur permettant d'explorer, puis expérimenter. Pour le premier, cet aspect systémique a été introduit par un système de jauges (santé, éducation, plaisir, etc.). Les actions de l'apprenant-joueur ont des effets sur les jauges, et celles-ci déterminent les conditions de victoire ou de défaite. Le second JS (*Jeu de l'Amazonie*) comporte une petite simulation qui est interrogée à chaque tour par les actions du joueur. Et comme dans l'exemple précédent, ces actions (planter, déforester, replanter, construire des logements, une usine, une école, etc.) ont un impact sur les variables (pas toujours rendues visibles aux apprenants-joueurs) et déterminent à la fois les actions futures possibles, mais aussi les conditions de victoire ou de défaite dans le JS.

Quatre des huit JS produits par les étudiants journalistes n'ont pas d'aspect systémique, même si deux d'entre eux contiennent des « jouets » (pour répondre aux demandes de l'enseignant). Dans les deux JS narratifs (*Rentabilise Neymar* et *Garde la pêche*), les étudiants sont parvenus à contourner les contraintes des outils auteurs (avec des pions réels, à comptabiliser à la fin de la partie) pour introduire un aspect systémique fondé sur des jauges. Les deux autres JS (*Enfance dans le monde* et *Jeu de*

*l'Amazonie*) s'approchent plus encore de la notion de jouet selon Ryan *et al.*, en intégrant un système avec lequel il est possible de jouer pour en expérimenter le fonctionnement et les limites. Dans ces deux derniers JS, les *playtests* ont montré que des apprenants-joueurs exploraient et expérimentaient le jouet systémique sans nécessairement suivre les buts des JS.

Nous avons constaté que les outils auteurs les plus polyvalents étaient très difficiles à prendre en main par des étudiants peu familiers des outils numériques. C'est ce qui a conduit certains groupes, après avoir fait un JS non numérique avec des aspects très systémiques, à se résoudre à réaliser un JS numérique non systémique, conçu avec des outils plus simples à appréhender comme *Twinery* ou *Klynt*, mais qui les ont contraints dans cette voie.

#### **4.2.3. Protocole pour les étudiants en *level design***

Nous avons introduit plusieurs différences dans l'étude menée avec les étudiants *level designers*, par rapport au protocole suivi avec les étudiants journalistes. À la place des travaux de Bogost *et al.* (2012) sur les *newsgames*, une réflexion a été menée sur *The Lens of the Toy* de Schell (2008) avec un travail d'identification de bons exemples de jouets dans les jeux vidéo populaires. De plus, le domaine a été imposé : les étudiants devaient concevoir un JS de niveau collègue sur le système immunitaire.

Au contraire des étudiants journalistes, les étudiants en *level design* étaient pour la plupart familiers du numérique et de la programmation et se sont facilement approprié l'outil auteur *Construct* à partir duquel tous les projets sauf un ont été menés.

#### **4.2.4. Productions des étudiants *level designers***

Les travaux des groupes d'étudiants *level designers* sont présentés dans le Tableau 5.

**Tableau 5 • Synthèse des JS produits par les étudiants  
level designers lors de l'étude 2**

Nom du JS (tous conçus avec Construct, sauf *)	Genre de JS choisi	Principales mécaniques	Méthode d'apprentissage visée
<i>Deficiencie</i>	<i>Tower defense</i>	Association de défenses à des vagues de microbes	Exploration + expérimentation
<i>Heal and Help *</i>	Jeu de plateau	Choisir des défenses pour contrer des événements infectieux	Exploration + expérimentation
<i>Immune Defense</i>	<i>Tower defense</i>	Association de défenses à des vagues de microbe	informations textuelles + entraînement
<i>Immunis</i>	Jeu de tir	Poursuivre et détruire les microbes ciblés par anticorps	informations textuelles + entraînement
<i>Infiltrhex</i>	Jeu de stratégie au tour par tour	À chaque tour, choisir des actions pour les microbes afin de traverser les défenses	Exploration + expérimentation

Nous observons que les JS réalisés appartiennent à des genres assez proches : le genre *tower defense* est assez proche du genre *jeu de stratégie au tour par tour*.

Nous distinguons néanmoins deux familles de choix de méthodes d'apprentissage visées par les étudiants, indépendantes des genres :

- d'une part, les JS dans lesquels il est prévu que l'apprentissage se fasse par l'apport d'informations (tutoriels) et son renforcement est prévu par l'entraînement ;
- d'autre part, ceux dans lesquels il est prévu que l'apprentissage se fasse en explorant, puis en expérimentant différentes solutions.

Nous avons constaté que l'ensemble des JS proposés ont des aspects systémiques, et nous pensons que cela est lié à trois causes :

- en premier lieu, la nature du sujet imposé aux étudiants (système immunitaire) ;
- en second lieu, l'absence de la barrière technique de la maîtrise de l'outil auteur générique ;

– en dernier lieu, la culture vidéoludique des étudiants, associée au travail d'inventaire qui a été fait de systèmes et des jouets dans les jeux vidéo populaires.

Pendant, seuls les JS qui proposent de l'exploration semblent être vraiment structurés autour d'un jouet au sens de Schell et de Ryan *et al.* En effet, ils sont les seuls à proposer des artefacts que les candidats aux *playtests* ont pris plaisir à manipuler indépendamment des instructions et des buts donnés dans les JS (Ryan *et al.*, 2012 ; Schell, 2008).

#### **4.2.5. Conclusion de la deuxième étude**

Nous avons tiré trois observations principales de l'ensemble de l'étude 2 :

- (1) L'introduction du concept de jauge semble faciliter la conception d'artefacts systémiques (Bogost *et al.*, 2012), mais pas forcément celle de jouets (Ryan *et al.*, 2012) ;
- (2) Le manque de maîtrise des outils auteurs polyvalents (*Construct*, *Gdevelop*) est un obstacle à l'implémentation de jouets dans les prototypes ;
- (3) Les sujets qui paraissent plus systémiques facilitent la création de jouets dans le jeu sérieux qui les traite. Donc, il nous faut proposer une méthode d'identification des aspects systémiques pour tous les types de sujets – étapes 1 et 2 de Ryan *et al.* (2012).

### **4.3. Troisième étude : avec une méthode de conception détaillée du jouet**

Cette troisième étude a impliqué 23 étudiants de DU de level design et 11 étudiants en M2 de journalisme.

#### **4.3.1. Protocole suivi**

Cette fois-ci, les deux formations ont été mélangées et les étudiants répartis équitablement sur trois projets de JS avec un thème imposé : L'Europe politique. Nous avons été convaincus d'instaurer ce mélange par le changement de stratégie de certains étudiants journalistes lors de la deuxième étude : ils avaient d'abord conçu des prototypes de JS non numériques, plutôt systémiques. Puis, finalement, ils avaient décidé de les abandonner face aux problèmes d'implémentation dans les outils auteurs (*Construct*, *Gdevelop*).

Comme nous avons (volontairement) imposé un thème qui pouvait sembler peu systémique aux étudiants, nous leur avons aussi fourni une méthode détaillée de conception de jouet, fondée d'une part sur Ryan *et al.*

(2012) articulée avec les 6 facettes (Marne, 2014) et, d'autre part, co-construite avec les précédents participants à partir des difficultés rencontrées lors des deux premières études. Cette méthode décrit le travail qui doit être opéré dans les facettes de la *Simulation du domaine* et des *Interactions avec la simulation*.

Le Tableau 6 contient le texte de l'une des fiches méthodologiques de synthèse fournies aux étudiants. En fournissant ces fiches aux étudiants, nous avons également l'ambition de tester si leur contenu pouvait fonder des patrons de conception pour l'élaboration de jeux sérieux articulés sur des jouets systémiques.

**Tableau 6 • Récapitulatif de la méthode de conception des jouets proposée aux étudiants**

Facette n° 2 : Concevoir la simulation	Facette n° 3 : Construire les interactions
<ol style="list-style-type: none"> <li>1. <b>Choisir</b> un sujet systémique</li> <li>2. Décrire et <b>modéliser</b> les systèmes               <ol style="list-style-type: none"> <li>1. Identifier les <b>paramètres</b> du système → <b>jauges</b></li> <li>2. Identifier les <b>événements</b> qui font varier les paramètres</li> <li>3. Identifier comment les paramètres <b>interagissent</b> entre eux</li> <li>4. Identifier les <b>règles</b> qui régissent le système</li> <li>5. <b>Simplifier</b> en se concentrant sur ce que l'on veut faire comprendre aux utilisateurs/trices</li> </ol> </li> <li>3. <b>Programmer</b> la simulation des systèmes               <ul style="list-style-type: none"> <li>◦ Système à base d'équations</li> <li>◦ Système à base de règles</li> <li>◦ Ou les deux</li> </ul> </li> <li>4. <b>Tester</b> la simulation pour vérifier...               <ul style="list-style-type: none"> <li>◦ Sa fiabilité</li> <li>◦ Sa simplicité</li> </ul> </li> </ol>	<ol style="list-style-type: none"> <li>1. Identifier comment le <b>hasard</b> peut jouer dans le système</li> <li>2. Imaginer une <b>métaphore</b> <ol style="list-style-type: none"> <li>1. Imaginer plusieurs <b>interactions</b> potentielles pour chaque paramètre</li> <li>2. Imaginer plusieurs <i>feedbacks</i> possibles pour chaque changement du système</li> <li>3. <b>Coordonner</b> interactions et <i>feedbacks</i></li> <li>4. Imaginer un environnement et des objets pour « <b>emballer</b> » les interactions et les <i>feedbacks</i></li> </ol> </li> <li>3. <b>Vérifier</b> que les interactions+<i>feedbacks</i> permettent...               <ol style="list-style-type: none"> <li>1. D'explorer le jouet pour essayer de comprendre le système : <b>explorer</b> les règles du système</li> <li>2. D'<b>expérimenter</b> les règles que les utilisateurs/trices ont comprises</li> </ol> </li> <li>4. <b>Tester</b> : est-il intéressant de manipuler le jouet, sans but et sans contrainte ?</li> </ol>



### 4.3.3. Productions des étudiants

Le Tableau 7 synthétise les JS produits lors de cette étude 3.

**Tableau 7 • Synthèse des JS produits lors de l'étude 3**

Nom du JS	Genre	Ambition	Principales mécaniques	Méthode d'apprentissage visée
<i>Enter the EU</i>	Simulation	Comprendre le processus d'entrée d'un pays dans l'UE	Choix binaires à chaque tour pour influencer des jauges	Exploration + expérimentation
<i>Élections européennes</i>	Succession de QCM	Comprendre la constitution de listes électorales	Choisir ou non les cartes qui sont présentées	Exploration + textes de remédiation
<i>Procédure législative</i>	Simulation	Comprendre l'élaboration des lois dans l'UE	Choix binaires à chaque tour pour influencer des jauges	Exploration + expérimentation

Nous avons fait deux observations principales. D'une part, les outils auteurs n'ont pas été un obstacle technique : les groupes étaient peuplés de suffisamment d'étudiants connaisseurs, si bien que le groupe *Élections européennes* a même choisi un outil auteur plus complexe que *Construct: Unity 3D*.

D'autre part, nous avons observé que pour deux des trois projets, les étudiants se sont emparés des éléments méthodologiques pour identifier ce qui faisait système dans leurs sujets afin de concevoir des simulations autour de jauges. Pour ces deux JS, l'interface choisie est celle de *Reigns* (Alliot *et al.*, 2016) qui propose un moyen simple de faire une interface ludique à une simulation multifactorielle. La Figure 1 est une capture d'écran de l'interface de ce jeu. Son principe est de présenter au joueur une situation décrite sommairement sur une carte. Une question fermée est posée au joueur à laquelle il ne peut répondre que positivement en glissant la carte vers la droite (mouvement de « *swipe* » sur les appareils mobiles) ou négativement en glissant la carte vers la gauche. En fonction de ce choix, les 4 jauges présentes en haut de l'écran (influence sur le clergé, influence sur la population, influence sur l'armée et budget) sont impactées positivement ou négativement. Puis, une nouvelle carte/situation/question est proposée au joueur. Tant que ces jauges n'atteignent pas zéro, la partie peut continuer.



Figure 1 • Capture d'écran du jeu *Reigns*, montrant l'interface

La figure 2 montre un exemple de la manière dont les étudiants se sont appropriés cette interface dans le JS *Enter the EU*: cette fois-ci ce sont des synthèses de textes de loi qui sont proposées à chaque tour, avec la possibilité de les présenter au parlement (tampon vert) ou de les rejeter (tampon rouge). Ce choix influence les jauges « Économie », « Social », « Politique », « Écologie ». Ce JS présente d'autres différences (ajout de conditions de victoire supplémentaires, d'événements plus marqués que dans *Reigns*, etc.).



Figure 2 • Capture d'écran du JS *Enter the EU* s'inspirant de *Reigns*

Les étudiants des deux groupes qui ont fait le choix de l'interface de *Reigns* ont pu constater à l'occasion des *playtests* que des participants exploraient et expérimentaient en *jouant* avec la simulation pour inférer la compréhension des systèmes, parfois sans poursuivre les buts du JS, c'est-à-dire exactement comme pour des jouets au sens de Ryan *et al.* (2012).

Des observations menées au cours de ces trois études, nous retenons de nombreux enseignements, dont les principaux sont présentés dans la partie suivante.

## **5. Principaux enseignements des trois études**

Nous revenons ici sur les deux principaux enseignements de ces trois études, dont nous pensons qu'ils doivent fonder de prochains travaux de recherche : d'une part, l'identification de trois verrous à la conception de JS fondés sur des jouets systémiques et, d'autre part, la proposition d'un patron de conception inspiré de l'interface de *Reigns*, dont le but est d'aider à lever un des verrous ou plusieurs d'entre eux.

### **5.1. Build the Toy First : 3 verrous principaux à la mise en œuvre**

Dans sa « *Lens of the Toy* » (2008), Schell conseille, pour faire ce qu'il appelle un bon jeu, de commencer par construire le jouet qui le fonde : « *Build the Toy First* ». Ce principe est repris dans les étapes proposées par Ryan *et al.* que nous avons décrites dans la section 2. Nous avons également associé ces étapes à 5 des 6 facettes de la conception des JS. Et, notamment, nous avons mis en évidence que la facette de la *Simulation du domaine* (facette 2) et la facette des *Interactions avec la simulation* (facette 3) étaient celles qui décrivaient la conception du jouet. Nous avons donc cherché à comprendre si cette association de méthodes était efficace pour construire, en premier, un jouet systémique dans le but de fonder un JS.

Selon nous, ces trois études confirment que la méthode de Ryan *et al.* associée aux 6 facettes est efficace à trois conditions. Les concepteurs doivent être capables de :

- (1) repérer les aspects systémiques du domaine que l'on cherche à enseigner ;
- (2) identifier et réaliser une interface claire et ludique pour permettre à l'apprenant-joueur de piloter cette simulation ;
- (3) maîtriser un outil auteur permettant d'intégrer cette simulation et son interface.

Au fil de nos études, nous avons essayé de co-construire des éléments méthodologiques complémentaires pour permettre aux concepteurs de lever ces trois verrous. Pour le premier d'entre eux, en nous appuyant sur les travaux de Ryan *et al.* (2012), sur ceux de Dormans (2012), sur nos propres travaux relatifs aux 6 facettes (Marne, 2014), et sur les travaux des participants, nous avons proposé des éléments de patrons de conception présentés dans le Tableau 6. Compte tenu des observations faites dans notre troisième étude, nous considérons que ces éléments de patrons, bien que trop rudimentaires actuellement, sont une bonne piste pour aider des concepteurs à envisager de commencer leurs JS par concevoir son jouet systémique, y compris pour des thématiques qui ne leur semblent pas systémiques au premier abord. Notamment, le concept éclairant pour les étudiants que nous avons observés a semblé être celui de *jaugé*.

En ce qui concerne le deuxième verrou, nous détaillons dans la sous-section suivante un patron de conception qui, en s'inspirant de l'interface de *Reigns*, tente d'apporter une aide méthodologique concrète.

Enfin, en ce qui concerne les outils auteurs, nous avons clairement observé l'influence que peuvent avoir ces outils sur le produit final. Par exemple, si techniquement il est possible avec des outils destinés à la narration interactive comme *Twinery* et *Klynt* de faire des jeux vidéo ou des JS avec des aspects systémiques<sup>5</sup>, la tâche est rendue ardue et surtout le fonctionnement de ces outils oriente grandement les concepteurs vers une narration linéaire branchée. C'est pourquoi nous pensons qu'il faut soit privilégier les outils auteurs plus génériques (*Gdevelopp*, *Construct*, *Unity 3D*) auprès des concepteurs capables de les manipuler (bonnes connaissances du numérique et compétences en algorithmique, voire en programmation), soit proposer de nouveaux outils auteurs, centrés sur la simulation et son interface, destinés à un public n'ayant pas ces prérequis. Le patron de conception présenté dans la section suivante pourrait, par son implémentation, servir de fondement à un outil auteur de ce type.

---

<sup>5</sup> Il existe sur le web de nombreux tutoriels comme (Greyelf, 2017) qui montrent étape par étape comment parvenir à incorporer un fonctionnement systémique dans ces outils de narration branchée.

## **5.2. Un patron complet et structuré, inspiré de *Reigns***

L'un des enseignements concrets des trois études précédentes que nous testons actuellement est la construction d'un patron de conception directement inspiré de l'interface du jeu vidéo *Reigns*, telle qu'elle a été utilisée par les étudiants de la troisième étude.

Ce patron reprend le formalisme décrit par Meszaros et Doble (1997), inspiré par Alexander *et al.* (1977) et déjà utilisé dans le contexte des EIAH (Delozanne *et al.*, 2007 ; Marne, 2014). Nous en présentons un extrait dans la sous-section 5.2.1, centré sur le problème posé et la solution qu'on peut y apporter. Le titre de la sous-section est le titre du patron de conception. La sous-section 5.2.2 qui suit cet extrait est un commentaire sur cet extrait.

### **5.2.1. Modalités et temporalités de l'interaction avec la simulation [Extrait du patron de conception]**

#### **Problème**

Comment construire une interface à cette simulation qui puisse être utilisée comme un jouet ou un micromonde, c'est-à-dire, dont l'exploration et l'expérimentation est à la fois ludique et instructive ?

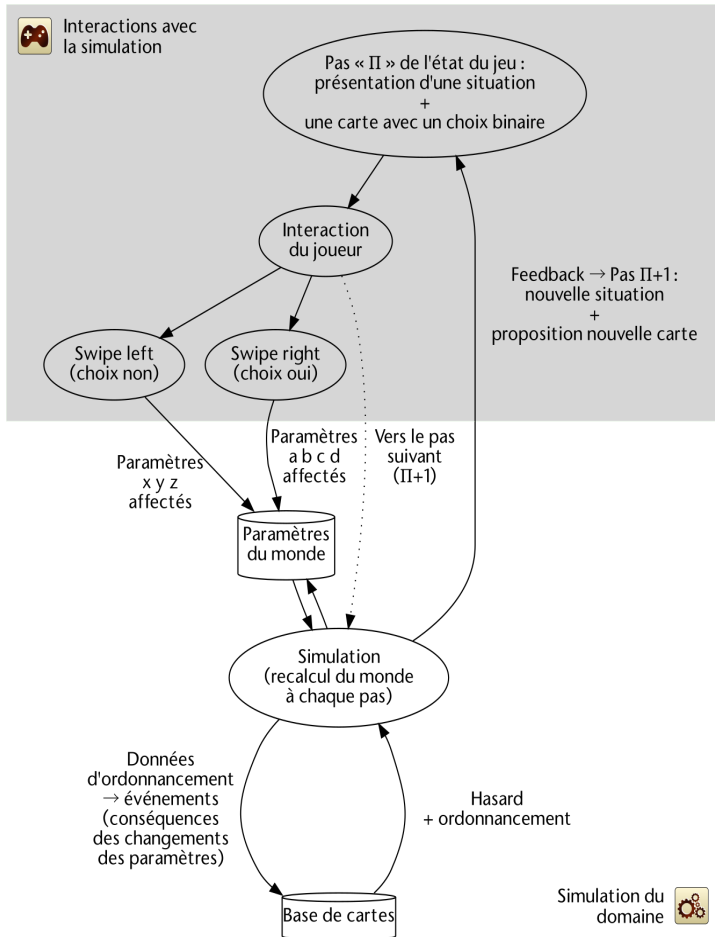
#### **Solution**

En vous inspirant du jeu vidéo *Reigns*, vous pouvez facilement créer une interface ludique avec les paramètres de la simulation, qui soit simple à concevoir et ludique pour les apprenants-joueurs.

*Reigns* est un jeu de gestion, où vous incarnez un roi. Le jeu est piloté par une simulation dans laquelle 4 paramètres principaux sont visibles sous forme de jauges (les relations avec le clergé, la population, l'armée et le budget).

L'interface choisie dans *Reigns* pour agir sur ces paramètres et donner du *feedback* au joueur est inspirée par les applications mobiles de rencontre comme *Tinder* : à chaque tour, une situation est présentée sur une carte au joueur. Elle s'accompagne d'une question. Si le joueur *swipe* à droite, c'est qu'il répond *oui* à la question, et *non* s'il *swipe* à gauche.

Le fonctionnement est détaillé dans la figure suivante.



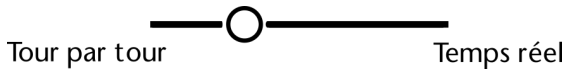
### Modélisation de la relation entre la simulation et son interface dans un jeu de type *Reigns*

Les tours de jeu sont appelés des « pas ». Le haut de la figure présente les moyens que l'on donne au joueur pour interagir avec les paramètres de la simulation (présentation du monde simulé, actions possibles, *feedbacks*). Le bas de la figure présente la façon dont la simulation recalcul le monde en fonction de ces interactions. a, b, c, d, x, y, z désignent des exemples de paramètres qui peuvent être modifiés par l'action du joueur. Entre chaque pas, la simulation recalcul l'état du monde, en tenant compte des paramètres modifiés par le joueur. Elle recalcul l'ensemble des paramètres, mais aussi les données temporelles (ordonnancement) qui peuvent être liées au pas courant, à l'action du joueur ou au hasard. Le recalcul du monde

permet de piocher ou de générer une nouvelle situation et une nouvelle question qui seront présentées au pas suivant par une nouvelle carte.

Vous pouvez réutiliser ce type d'interaction tel quel dans votre jeu sérieux, c'est-à-dire avec une proposition binaire d'interactions oui/non et des *pas* distingués par la présentation de situations/questions. Mais vous pouvez aussi faire varier deux aspects. [...]

**La temporalité (durée du pas).** Dans *Reigns*, le joueur peut prendre son temps pour décider de son action (répondre oui ou non) à chaque *pas*. À la place, vous pouvez aussi proposer un *pas* qui n'est pas temporisé par l'action du joueur, c'est-à-dire que si le joueur n'agit pas au bout d'un temps défini, le *pas* suivant est déclenché. Vous avez ainsi la possibilité de faire varier la durée du *pas* avec, à un extrême, le *pas* terminé par l'action du joueur (tour par tour) comme dans *Reigns*, et, à un autre extrême, le *pas* exécuté automatiquement aussi vite que possible (temps réel) comme dans *Ancestorsn: The Humankind Odyssey*. À vous de choisir où vous placez le curseur.



**La modalité (type d'actions autorisées).** Dans *Reigns*, le joueur ne peut agir qu'en répondant oui ou non. À la place, vous pouvez proposer plus de modalités d'interaction, par exemple en autorisant plus de types de réponses, ou d'autres types d'actions que de répondre à des questions. Dans le processus de conception, il est judicieux d'augmenter progressivement le nombre de modalités disponibles : en partant du premier extrême qui est celui de *Reigns* avec la modalité la plus rudimentaire (choix binaire), en allant vers l'autre extrême dans lequel les choix d'actions sont très nombreux et diversifiés (choix multimodaux) comme dans *Crusader Kings*. À vous de choisir progressivement où arrêter le curseur.



### **5.2.2. À propos de ce patron de conception**

L'identification du patron de conception présenté dans la sous-section précédente était évidente dans la 3<sup>e</sup> étape de l'expérimentation, puisque le jeu *Reigns* a été explicitement pris comme modèle par les étudiants. Mais, nous l'avons aussi repéré dans la deuxième étude. Dans les JS *Enfance dans le monde* et *Jeu de l'Amazonie* décrits dans le Tableau 3 et le Tableau 4 les étudiantes impliquées avaient déjà construit leurs interactions avec la simulation en utilisant certains principes de ce patron : discriminer les paramètres, la temporalité des actions du joueur, et les modalités de ces actions. Et même dans la première étude, dans le JS au tour par tour *Citizens of Nowhere*, décrit dans le Tableau 1 et le Tableau 2, les étudiants avaient aussi travaillé la question de la temporalité et des modalités d'action des apprenants-joueurs.

Aujourd'hui encore, nous observons ce patron de conception dans la suite de l'expérimentation (itération 4) : les étudiants impliqués (connaissant *Reigns* et les travaux de leurs prédécesseurs) ont voulu réutiliser des interfaces similaires. Il est difficile de tirer des conclusions ou de discuter davantage sur cette expérimentation encore en cours.

Nous pensons que ce patron de conception pourrait permettre à des concepteurs de déterminer plus facilement la bonne interface avec la simulation du domaine. Mais, nous pensons aussi que ce patron pourrait clarifier dans l'esprit des concepteurs le rôle que peuvent jouer la simulation et son interface dans la conception d'un jouet systémique/micromonde. Ces nouvelles hypothèses sont à tester dans de futures études.

D'autres travaux sont à mener sur le formalisme des patrons de conception présentés ici, afin de parvenir à en produire une implémentation sous forme d'outil auteur. En effet, un outil auteur inspiré du patron « Modalités et temporalités de l'interaction avec la simulation » semble réalisable, et pourrait, comme le patron lui-même, être un outil de conception intéressant pour des JS fondés sur des jouets systémiques ou des micromondes.

## **6. Conclusion**

Dans la perspective de fournir des Jeux Sérieux (JS) permettant un apprentissage par l'exploration et l'expérimentation grâce à des jouets/micromondes très systémiques, nous avons choisi de mener trois études de terrain en situation écologique. Inspirées de la recherche



collaborative orientée par la conception, elles ont été menées auprès d'une centaine d'étudiants qui ont participé à la conception de 23 prototypes de JS. Les objectifs de ces études étaient, d'une part, de mettre à l'épreuve la méthode de conception proposée par Ryan *et al.* (2012) et, d'autre part, d'extraire des éléments de patrons de conception pour les JS centrés sur des jouets/micromondes.

Nous avons observé que, pour que les participants appréhendent correctement la notion du jouet systémique, comme un micromonde, il est nécessaire d'aller au-delà de la définition et des éléments méthodologiques proposés par Schell dans *The Lens of the Toy* (2008). Ainsi, nous montrons que les trois étapes de conception de jouets proposées par Ryan *et al.* (2012) associées aux 6 facettes de la conception des jeux sérieux présentées dans (Marne *et al.*, 2011) ont aidé les participants à intégrer et concevoir leurs JS autour de jouets systémiques. Toutefois, nous avons aussi identifié trois verrous à lever dans cette méthode : (1) l'identification des éléments de simulation dans les sujets qui paraissent les moins systémiques ; (2) la conception d'interactions ludiques pour cette simulation ; (3) l'influence parfois négative des outils auteurs sur la construction de JS à partir de jouets/micromondes.

Pour les deux premiers verrous, nous proposons des patrons de conception, dont le patron « Modalités et temporalités de l'interaction avec la simulation », qui restent à raffiner ou à tester. Quant au troisième verrou, nous proposons de diriger les concepteurs compétents vers les outils auteurs génériques (du type de *Unity 3D* ou de *Construct*) et nous proposons d'orienter nos futures recherches vers l'implémentation, sous forme d'outil auteur simplifié, de ce même patron de conception pour permettre aux concepteurs non-informaticiens de développer des JS fondés sur des jouets/micromondes.

## **REMERCIEMENTS**

L'auteur remercie le LABEX ASLAN (ANR-10-LABX-0081) de l'Université de Lyon pour son soutien financier dans le cadre du programme « Investissements d'Avenir » (ANR-11-IDEX-0007) de l'État français géré par l'Agence Nationale de la Recherche (ANR).

## **RÉFÉRENCES**

Alexander, C., Ishikawa, S. et Silverstein, M. (1977). *A pattern language: Towns, buildings, construction*. Oxford University Press.

Alliot, F., Murakami, M., Vreeland, R. et Lugo, M. (2016). Reigns [jeu vidéo]. Nerial, Devolver Digital. <https://www.devolverdigital.com/games/reigns>

Bogost, I. (2007). *Persuasive games: The expressive power of videogames*. MIT Press.

Bogost, I. (2013). Exploitationware. Dans R. Colby, M. S. S. Johnson et R. S. Colby (dir.), *Rhetoric/Composition/Play through video games: Reshaping theory and practice of writing* (p. 139-147). Palgrave Macmillan US.

Bogost, I., Ferrari, S. et Schweizer, B. (2012). *Newsgames: Journalism at play*. MIT Press.

Brougère, G. (2013). Les jouets et la rhétorique de l'éducation. *Sociographe*, 1(41), 81-90.

Callan, R. C., Bauer, K. N. et Landers, R. N. (2015). How to avoid the dark side of gamification: Ten business scenarios and their unintended consequences. Dans T. Reiners, L. C. Wood, *Gamification in education and business* (p. 553-568). Springer International.

Crawford, C. (1984). *The art of computer game design*. Osborne/McGraw-Hill.

Davis, J. P., Steury, K. et Pagulayan, R. (2005). A survey method for assessing perceptions of a game: The consumer playtest in game design. *Game Studies*, 5(1).

Deci, E. L., Koestner, R. et Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological bulletin*, 125(6), 627.

Delozanne, E., Le Calvez, F., Merceron, A. et Labat, J.-M. (2007). Design patterns en EIAH : vers un langage de Patterns pour l'évaluation des apprenants. *Sticef*, 14, 45-80.

Dormans, J. (2012). *Engineering emergence: Applied theory for game design*. Amsterdam Creative Commons.

Drachen, A., Mirza-Babaei, P. et Nacke, L. E. (dir.) (2018). *Games user research*. Oxford University Press.

Garris, R., Ahlers, R. et Driskell, J. E. (2002). Games, motivation, and learning: A research and practice model. *Simulation & gaming*, 33(4), 441-467.

Gee, J. P. (2009). Deep learning properties of good digital games: How far can they go? Dans U. Ritterfeld, M. Cody et P. Vorderer (dir.), *Serious games: mechanisms and effects* (chap. 5, p. 67-82). Routledge. <https://doi.org/10.4324/9780203891650>

Greyelf, (2017, 5 juillet). Best answer given to the online forum question How can I implement a visual Health Bar in Harlowe 2.0.1?. Twine Q&A. <https://twinery.org/questions/351/how-can-i-implement-a-visual-health-bar-in-harlowe-2-0-1?show=364#a364>

Hugon, M.-A. et Seibel, C. (dir.). (1988). *Recherches impliquées, recherches action : Le cas de l'éducation*. De Boeck.

James, A., Clauss, G., Curé-heaton, J., Gizard, D., Madinier, B., Piron, T., Soupama, É., Vesir, M. et Zerarka, A. (2016). Isis the end? [jeu vidéo] <http://www.isistheend.com/>

Keller, J. M. (1995). Motivation in cyber learning environments. *International Journal of Educational Telecommunications*, 1(1), 7-30.

Malone, T. W. et Lepper, M. R. (1987). Making learning fun: A taxonomy of intrinsic motivations for learning. Dans R. E. Snow et M. J. Farr (dir.), *Aptitude, learning and instruction: Vol. 3. Conative and affective process analyses* (chap. 10, p. 222-253). Hillsdale, NJ : Erlbaum.

Marfisi-Schottman, I. (2012). *Méthodologie, modèles et outils pour la conception de Learning Games* [thèse de doctorat, INSA de Lyon, France]. <http://theses.insa-lyon.fr/publication/2012ISAL0103/these.pdf>

Marfisi-Schottman, I. et Piau-Toffolon, C. (2015). Extraire et réutiliser des patrons de conception à partir de Learning Games existants. Dans S. George, G. Molinari, C. Cherkaoui, D. Mammass et L. Oubahssi (dir.), *Actes de l'atelier Méthodologies de conception collaborative des EIAH: vers des approches pluridisciplinaires (EIAH 2015)* (p. 24-35). ATIEF. [http://www.atief.fr/sitesConf/eiah2015/uploads/EIAH-2015\\_Atelier\\_Conception\\_Actes.pdf](http://www.atief.fr/sitesConf/eiah2015/uploads/EIAH-2015_Atelier_Conception_Actes.pdf)

Marne, B. (2014). *Modèles et outils pour la conception de jeux sérieux : une approche meta-design* [thèse de doctorat en informatique, Université Pierre et Marie Curie, Paris, France]. <https://tel.archives-ouvertes.fr/tel-01134701>

Marne, B., Huynh-Kim-Bang, B. et Labat, J.-M. (2011). Articuler motivation et apprentissage grâce aux facettes du jeu sérieux. Dans M. Bétrancourt, C. Depover, V. Luengo, B. De Lièvre et G. Temperman (dir.), *Actes de la conférence Environnements Informatiques pour l'Apprentissage Humain (EIAH 2011)* (p. 69-80). Université de Mons.

Meszaros, G. et Doble, J. (1997). A pattern language for pattern writing. Dans R. C. Martin, D. Riehle et F. Buschmann, *Pattern languages of program design-3* (p. 529-574). Addison-Wesley Longman.

Moreno-Ger, P., Martinez-Ortiz, I. et Fernández-Manjón, B. (2005). The E-Game Project: Facilitating the development of educational adventure games. Dans D. Kinshuk, P. Sampson, L. Isaías, P. Rodrigues et P. Barbosa (dir.), *Proceedings of the IADIS International Conference on Cognition and exploratory learning in digital age (CELDA 2005)* (p. 353-358). IADIS.

Papert, S. (1987). Microworlds: Transforming education. *Artificial intelligence and education*, 1, 79-94.

Peirce, C. S. (1902). *Minute Logic: Chap. I. Intended Characters of this Treatise*. MS [R] 425.

Premsky, M. (2004). *Digital game-based learning*. McGraw-Hill.

Ryan, M., Costello, B. et Stapleton, A. (2012, octobre). *Deep learning games through the lens of the toy* [communication]. Meaningful Play 2012 Conference, East Lansing, MI, USA.

Sanchez, É. et Monod-Ansaldi, R. (2015). Recherche collaborative orientée par la conception. Un paradigme méthodologique pour prendre en compte la complexité des situations d'enseignement-apprentissage. *Education & didactique*, 9(2), 73-94.

Schell, J. (2008). *The art of game design: A book of lenses*. Morgan Kaufmann.

Wang, F. et Hannafin, M. J. (2005). Design-based research and technology-enhanced learning environments. *Educational Technology Research and Development*, 53(4), 5-23.

Wouters, P., Van Nimwegen, C., Van Oostendorp, H. et Van Der Spek, E. D. (2013). A meta-analysis of the cognitive and motivational effects of serious games. *Journal of Educational Psychology*, 105(2), 249-265. <https://doi.org/10.1037/a0031311>





## Comités

---

---

### Rédacteur en chef

Sébastien GEORGE • LIUM, Le Mans Université, Laval

### Comité de rédaction

Georges-Louis BARON • EDA, Université de Paris

Monique BARON • LIP6, Sorbonne Université

Laetitia BOULC'H • EDA, Université de Paris

Éric BRUILLARD • EDA, Université de Paris

Pierre-André CARON • CIREL, Université de Lille

Christophe DESPRÈS • LIUM, Le Mans Université

Nour EL MAWAS • CIREL, Université de Lille

Sébastien GEORGE • LIUM, Le Mans Université, Laval

Monique GRANDBASTIEN • LORIA, Université de Lorraine

Richard HOTTE • LICEF, Télé-université, Université du Québec, Montréal,  
Canada

Pierre JACOBONI • LIUM, Le Mans Université

Élise LAVOUÉ • LIRIS, Université Jean Moulin Lyon 3

Vanda LUENGO • LIP6, Sorbonne Université

Agathe MERCERON • Université de Berlin, Allemagne

Gaëlle MOLINARI • TECFA, Unidistance, Genève, Suisse

Chrysta PÉLISSIER • Praxiling, Université Montpellier 3

Jean-Luc RINAUDO • CIRNEF, Université de Rouen Normandie

### Comité de parrainage scientifique

Nicolas BALACHEFF • LIG, CNRS

Stefano CERRI • LIRMM, Université de Montpellier 2

Christian DEPOVER • Université de Mons, Belgique

Alain DERYCKE • TRIGONE, Université de Lille

Pierre DILLENBOURG • École polytechnique fédérale de Lausanne, Suisse

Claude FRASSON • Université de Montréal, Canada

Catherine GARBAY • LIG, CNRS

Gilles GAUTHIER • UQAM, Canada

Guy GOUARDÈRES • API, Université Paul Sabatier

Ulrich HOPPE • Université de Duisbourg, Allemagne

Jean-Marc LABAT • LIP6, Sorbonne Université  
Patrick MENDELSON • LSE, IUFM Académie de Grenoble  
Jean-François NICAUD • LIG, Université Grenoble Alpes  
Gilbert PAQUETTE • LICEF, Télé-université, Université du Québec,  
Montréal, Canada  
Janine ROGALSKI • DR CNRS honoraire, LDAR, Université de Paris  
Maria Felisa VERDEJO • Universidad nacional de educación a distancia,  
Espagne

### **Comité de lecture**

Michel ARNAUD • Université Paris Nanterre  
François-Xavier BERNARD • EDA, Université de Paris  
Mireille BÉTRANCOURT • TECFA, Université de Genève, Suisse  
Jacques BÉZIAT • CIRNEF, Université de Caen Normandie  
Bernard BLANDIN • CREF, Université Paris Nanterre et CESI  
François BOUCHET • LIP6, Sorbonne Université  
Julien BROISIN • IRT, Université de Toulouse Paul Sabatier  
Thibault CARRON • LIP6, Sorbonne Université et Université de Savoie  
Mont-Blanc  
Ullrich CARSTEN • EdTec Lab, DFKI GmbH, Sarrebrück, Allemagne  
Thierry CHANIER • LRL, Université Clermont Auvergne  
Ghislaine CHARTRON • CNAM, Paris  
Christophe CHOQUET • LIUM, Le Mans Université, Laval  
Philippe COTTIER • CREN, Université de Nantes  
Jacques CRINON • INSPÉ, Université Paris Est Créteil  
Bruno DE LIÈVRE • Université de Mons, Belgique  
Nicolas DELESTRE • LITIS, INSA de Rouen  
Élisabeth DELOZANNE • LIP6, Sorbonne Université  
Michel DESMARAIS • École polytechnique de Montréal, Canada  
Cyrille DESMOULINS • LIG, Université Grenoble Alpes  
Philippe DESSUS • LSE, Université Grenoble Alpes  
Angélique DIMITRACOPOULOU • LTEE, Université d'Égée, Grèce  
Béatrice DROT-DELANGÉ • ACTé, Université Clermont Auvergne  
Aude DUFRESNE • ESI, Université de Montréal, Canada  
Cédric FLUCKIGER • Théodile-CIREL, Université de Lille  
Serge GARLATTI • Lab-STICC, IMT Atlantique, Brest  
Jean-Marie GILLIOT • Lab-STICC, IMT Atlantique, Brest  
Brigitte GRUGEON • LDAR, INSPÉ, Université Paris Est Créteil  
Viviane GUÉRAUD • LIG, Université Grenoble Alpes  
Nicolas GUICHON • ICAR, Université Lumière Lyon 2

Nathalie GUIN • LIRIS, Université Lyon 1  
 France HENRI • LICEF, Télé-université, Université du Québec, Montréal, Canada  
 Pierre JARRAUD • FOAD, Sorbonne Université  
 Michelle JOAB • LIRMM, Université Montpellier 2  
 Céline JOIRON • MIS, Université de Picardie Jules Verne, Amiens  
 Mehdi KHANEBOUBI • STEF, ENS Paris-Saclay  
 Vassilis KOMIS • Université de Patras, Grèce  
 Thérèse LAFERRIÈRE • TACT, Université Laval, Canada  
 Françoise LE CALVEZ • LIP 6, Sorbonne Université  
 Marie LEFÈVRE • LIRIS, Université Lyon 1  
 Dominique LENNE • Heudiasyc, Université de Technologie de Compiègne  
 Pascal LEROUX • CREN, Le Mans Université  
 Paul LIBBRECHT • Leibniz Institute for Research and Information in Education, Allemagne  
 Cabral LIMA • Université Fédéral de Rio de Janeiro, Brésil  
 Domitile LOURDEAUX • Heudiasyc, Université de Technologie de Compiègne  
 Pascal MARQUET • LISEC, Université de Strasbourg  
 Jean-Charles MARTY • LIRIS, Université de Savoie  
 André MAYERS • Université de Sherbrooke, Canada  
 Christine MICHEL • TECHNÉ, Université de Poitiers  
 Roger NKAMBOU • GDAC, Université du Québec à Montréal, Canada  
 Thierry NODENOT • LIUPPA, Université de Pau et des Pays de l'Adour, Bayonne  
 Daniel PERAYA • TECFA, Université de Genève, Suisse  
 Yvan PETER • CRISTAL, Université de Lille  
 Julia PILET • LDAR, INSPÉ, Université Paris Est Créteil Val de Marne  
 Dominique PY • LIUM, Le Mans Université  
 Christophe REFFAY • ELLIAD, INSPÉ, Université de Franche-Comté  
 Éric SANCHEZ • TECFA, Université de Genève, Suisse  
 Karim SEHABA • LIRIS, Université Lyon 2  
 Nicolas SZILAS • TECFA, Université de Genève, Suisse  
 Pierre TCHOUNIKINE • LIG, Université Grenoble Alpes  
 André TRICOT • EPSYLON, Université Paul-Valéry Montpellier 3  
 Nicolas VAN LABEKE • Learning Sciences Research Institute, University of Nottingham, UK  
 Jean VANDERDONCKT • ISYS, Université catholique de Louvain, Belgique  
 Kalina YACEF • Université de Sydney, Australie

**En mémoire d'anciens membres des comités :**

Erik DUVAL • Université de Louvain, Belgique

Jacques PERRIAULT • Université Paris Nanterre

François VILLEMONTÉIX • CIREL, Université de Lille

**Nous remercions les personnes extérieures aux comités  
qui ont relu pour ce numéro :**

Jean-Marie BOILEVIN • CREAD, INSPÉ Bretagne

Denis BOUHINEAU • LIG, Université Grenoble Alpes

Jérôme ENEAU • CREAD, Université de Rennes 2

Nadine MANDRAN • LIG, Université Grenoble Alpes / CNRS

Sandra NOGRY • Paragraphe, Université de Cergy-Pontoise

Ecatérina PACURAR • CIREL, Université de Lille

Gaëtan TEMPERMAN • Université de Mons







ISBN 978-2-901384-04-5  
DOI: 10.23709/sticef.27.2