



HAL
open science

Évaluation formative et génération de feedbacks soutenues par la technologie

Franck Silvestre

► **To cite this version:**

Franck Silvestre. Évaluation formative et génération de feedbacks soutenues par la technologie. Environnements Informatiques pour l'Apprentissage Humain. Université Toulouse Capitole, 2024. tel-04567491

HAL Id: tel-04567491

<https://hal.science/tel-04567491v1>

Submitted on 3 May 2024

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

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



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0
International License

Mémoire

En vue de l'obtention de l'

Habilitation à Diriger des Recherches

Délivrée par l'Université Toulouse Capitole
Discipline Informatique

Présentée par Franck Silvestre
le 5 avril 2024

Évaluation formative et génération de feedbacks soutenues par la technologie

JURY

Philippe Dessus	Professeur, Université Grenoble Alpes	Rapporteur
Sébastien George	Professeur, Le Mans Université	Rapporteur
Élise Lavoué	Professeure, Université Jean-Moulin Lyon 3	Rapportrice
Frédéric Amblard	Professeur, Université Toulouse Capitole	Examineur
Grégoire Borst	Professeur, Université Paris Cité	Examineur
Julien Broisin	Maître de conférences HDR, Université Toulouse 3 Paul Sabatier	Examineur

École doctorale : Mathématiques, Informatique et Télécommunications de Toulouse
Unité de recherche : Institut de Recherche en Informatique de Toulouse – IRIT UMR 5505 CNRS

Remerciements

Mes premiers remerciements vont à mon premier compagnon de route en recherche : Julien Broisin. Merci pour tes soutiens indéfectibles sur ces sentiers exigeants de la recherche, qu'ils soient sur un plan scientifique ou personnel.

Je tiens à remercier les membres de l'équipe TALENT qui constituent le premier cercle dans lequel j'effectue ma recherche. Elle ne serait pas aussi passionnante, si cette équipe n'était pas aussi accueillante et riche par ses échanges et interactions. Merci à Mar pour les pauses déjeuner *brain storming*. Remerciements appuyés aux doctorants passés, présents et futurs que j'ai (eu) immense plaisir à co-encadrer : Rialy, Mika, Denis et Rémi. Ces thèses n'apporteraient pas autant de satisfaction sans mes collègues avec qui nous co-encadrons : Julien, Jean-Baptiste et Jean-Michel. Merci à Jean-François pour ses contributions scientifiques indispensables à nos travaux. Merci à John pour ses contributions indispensables à l'ingénierie de nos logiciels.

Merci à Frédéric Amblard d'avoir accepté sans hésitation d'être mon « garant » sur cette démarche d'obtention de l'HDR. Le soutien de Frédéric à l'Université Toulouse Capitole m'est très précieux. Merci à Marylène, Jean-François, Frédéric, Jean-Michel et Mar pour la relecture de ce manuscrit avant sa mise entre les mains des rapporteurs. Merci aux membres du jury qui ont accepté de rapporter ce document. Sébastien, Élise et Philippe, cela a été une grande joie d'apprendre que vous acceptiez. Je m'en suis senti honoré. Merci à Julien, à Frédéric et à Grégoire d'avoir accepté de prendre part à ce jury et de participer ainsi, avec les rapporteurs, à la construction de mes futures recherches.

La science en train de se faire, se fait au sein d'une communauté scientifique : je trouve la nôtre particulièrement riche en raison de sa nature pluri-inter-trans-disciplinaire. Merci à tous les membres de cette communauté pour les échanges, les apports, les idées. Remerciements appuyés à Nadine et à Christian pour un moment privilégié une après-midi d'EIAH 2023, quand nous patientions pour la visite de l'aquarium à Brest.

Je m'efforce de travailler toujours davantage avec les praticiens, les enseignants et les élèves/étudiants sur le terrain. Merci à Gilles et à Sylvie, de l'Académie de Nancy-Metz, merci à Estelle, merci aux collègues du collège d'Ayguesvives, merci à vous tous pour la confiance et l'enthousiasme. Merci à tous les enseignants et les élèves/étudiants qui contribuent, en utilisant nos logiciels, à nos travaux de recherche.

Et, au-delà de tous ceux qui contribuent directement à cette recherche, il y a tous ceux

qui y contribuent indirectement en participant, dans la sphère privée, à la construction d'un cadre de vie que je ne changerai pour rien au monde. Merci donc à la famille et aux amis. Je remercie intimement Marylène, mon épouse, et Thomas, mon fils, à qui je dois tant.

Résumé

Ce document présente une synthèse de mes travaux de recherche conduits au sein de l'Institut de Recherche en Informatique de Toulouse. Ces travaux s'inscrivent dans le champ large des recherches en Ingénierie des Environnements Informatiques pour l'Apprentissage Humain (IEIAH). Je m'intéresse depuis ma thèse débutée en 2012 à la manière dont la technologie peut venir en soutien de l'évaluation formative mise en œuvre en classe ou hors classe.

Les recherches sur les situations en classe ont conduit à des contributions portant sur les processus d'évaluation formative mobilisables en face-à-face (processus en N phases, processus à 2 votes) ainsi que dans la conception de systèmes interactifs facilitant la mise en œuvre de séquences pédagogiques mobilisant de tels processus (Tsaap-Notes, Elaastic). Ces recherches ont été réalisées dans le cadre de deux thèses : la mienne et celle de Rialy Andriamiseza, soutenue en 2022, que j'ai codirigée avec Julien Broisin.

Les recherches sur les situations hors classe ont débuté dans le cadre de ma thèse. Ces travaux ont conduit à des contributions sur la génération automatique de feedbacks pour les tests autocorrectifs. Ils se poursuivent aujourd'hui dans le cadre de la thèse de Denis Ollivier codirigée avec Julien Broisin et coencadrée avec Jean-Baptiste Raclet. Ces travaux en cours portent sur l'ingénierie des systèmes interactifs soutenant les activités de révisions des apprenants.

L'intersection de mes recherches avec la psychologie, les neurosciences et les sciences de l'éducation m'ont mené à une relecture des travaux réalisés ces 10 dernières années. Cette relecture m'a permis de formuler une nouvelle situation prototypique pour les recherches en IEIAH. Celle-ci est fondée sur l'étude des systèmes technologiques soutenant (respectivement limitant) les états cognitifs et émotionnels favorables (respectivement défavorables) à l'apprentissage humain. C'est dans ce cadre que sont formulés les axes du projet de recherche présenté dans ce document. Les problématiques abordées concernent l'adaptation des processus d'évaluation formative et la génération, par des agents humains ou artificiels, de feedbacks textuel personnalisé.

Mots clés Évaluation formative, feedback, systèmes interactifs, *learning analytics*, *design-based research*, méthodes mixtes, système 1 et système 2, évaluations par les pairs, intelligences artificielles génératives.

Abstract

This document summarizes my research work at the Institut de Recherche en Informatique de Toulouse. This work falls within the broad field of research in Technology-Enhanced Learning (TEL). Since starting my thesis in 2012, I've been interested in how technology can support formative assessment in and out of the classroom.

Research on classroom situations has led to contributions on formative assessment processes that can be mobilized face-to-face (N-phase process, 2-votes-based process), as well as in the design of interactive systems facilitating the implementation of pedagogical sequences mobilizing such processes (Tsaap-Notes, Elaastic). This research was carried out as part of two PhD thesis : my own and that of Rialy Andriamiseza, defended in 2022, which I co-directed with Julien Broisin.

Research on out-of-class situations began as part of my PhD thesis. This work led to contributions on the automatic generation of feedback for self-correcting tests. It continues today in the context of Denis Ollivier's thesis, co-directed with Julien Broisin and co-supervised with Jean-Baptiste Raclet. This ongoing work focuses on the engineering of interactive systems to support learners' revision activities.

The intersection of my research with psychology, neuroscience and educational science has led me to reconsider the work carried out over the last 10 years. This review has guided the formulation of a new prototypical situation for TEL research. This is based on the study of technological systems supporting (respectively limiting) cognitive and emotional states favorable (respectively unfavorable) to human learning. It is within this framework that the main lines of the research project presented in this document are formulated. The mentioned issues concern the adaptation of formative evaluation processes and the generation, by human or artificial agents, of personalized textual feedback.

Keywords Formative assessment, feedback, interactive systems, learning analytics, design-based research, mixed methods, system 1 and system 2, peer evaluations, generative artificial intelligences.

Table des matières

1	Introduction	1
1.1	Contexte général	1
1.2	Axes de recherche et problématiques	2
1.2.1	La technologie en soutien de l'évaluation formative en classe	3
1.2.2	La technologie en soutien de l'évaluation formative hors classe	4
1.3	Méthode de conduite de la recherche	4
1.3.1	Posture épistémologique	4
1.3.2	Méthode	6
1.3.3	Éthique de l'observation instrumentée	6
1.4	Organisation du manuscrit	7
I	Travaux de recherche	8
2	La technologie en soutien de l'évaluation formative en classe	9
2.1	État de l'art	10
2.1.1	Les notions clés	10
2.1.2	Les évaluations formatives et leur mise en œuvre dans la classe	11
2.1.3	Les évaluations formatives soutenues par la technologie	12
2.2	Problématiques et questions de recherche	14
2.2.1	Conception d'un système d'évaluation d'audience	14
2.2.2	Compréhension et amélioration des processus à 2 votes	16
2.3	Conception d'un système d'évaluation d'audience	17
2.3.1	Méthode de conduite de la recherche	17
2.3.2	Identification des processus d'évaluation d'audience	17
2.3.3	Évaluation et comparaison des processus d'évaluation d'audience	18
2.3.4	Le processus en N Phases	20
2.3.5	Tsaap-Notes	20
2.3.6	Expérimentations et résultats	24
2.3.7	Discussion et positionnement des travaux	26
2.3.8	Rétrospective	29

TABLE DES MATIÈRES

2.4	Compréhension et amélioration des processus à 2 votes	31
2.4.1	Méthode de conduite de la recherche	31
2.4.2	L'analyse de données exploratoire	32
2.4.3	Résultats et recommandations induites	35
2.4.4	Un nouveau processus à 2 votes	38
2.4.5	Expérimentation et résultats	41
2.4.6	Discussion et positionnement des travaux	44
2.4.7	Rétrospective	47
2.5	Conclusion	49
3	La technologie en soutien de l'évaluation formative hors classe	51
3.1	État de l'art	52
3.1.1	Les tests autocorrectifs	52
3.1.2	Les effets test et espacement	53
3.2	Problématiques et questions de recherche	53
3.2.1	L'automatisation de la construction de tests autocorrectifs	53
3.2.2	La conception d'un système de révision	55
3.3	La génération semi-automatique de tests autocorrectifs	55
3.3.1	Méthode de conduite de la recherche	55
3.3.2	Première itération	56
3.3.2.1	L'approche <i>Note as feedback</i>	56
3.3.2.2	Expérimentation et résultats	57
3.3.3	Seconde itération	60
3.3.3.1	L'approche <i>Rationale as feedback</i>	60
3.3.3.2	Expérimentation et résultats	61
3.3.4	Discussion et positionnement des travaux	62
3.4	La conception d'un système de révision	64
3.4.1	Méthode de conduite de la recherche	64
3.4.2	L'étude qualitative exploratoire	65
3.4.2.1	Participants et collecte des données	66
3.4.2.2	Résultats et hypothèses induites	67
3.4.2.3	L'application Konsolidation	69
3.4.3	Discussion et positionnement des travaux	69
3.4.4	Rétrospective	71
3.5	Conclusion	72
II	Le projet de recherche	73
4	Retour sur la situation prototypique d'une recherche en IEAH	74

TABLE DES MATIÈRES

4.1	De l'artefact à plusieurs instruments	74
4.2	Proposition d'une nouvelle situation prototypique	75
4.3	Elaastic passé au crible de la nouvelle situation prototypique	78
4.4	Conclusion	79
5	Programme de recherche	80
5.1	Ingénierie des EIAH	80
5.1.1	Soutiens à l'activation du système 2 et à l'inhibition du système 1	80
5.1.2	Soutien à la production de feedbacks personnalisés	82
5.1.3	Intégration des nouveaux axes de recherche	83
5.2	Méthode de conduite de la recherche	84
5.2.1	Soutien à la démarche expérimentale en contexte écologique	84
5.2.2	Formation et accompagnement des doctorants	85
5.3	Conclusion	87
III	Rapport d'activités	89
6	<i>Curriculum Vitae</i> résumé	90
6.1	Données personnelles	90
6.2	Parcours professionnel	90
6.3	Cursus universitaire	91
7	Activités de recherche	93
7.1	Présentation synthétique des thématiques de recherche	93
7.2	Encadrements scientifiques	94
7.3	Prix et distinctions	96
7.4	Diffusion du savoir, responsabilités et activités au sein des sociétés savantes ou associations	97
7.5	Organisation colloques, conférences, journées d'étude	99
7.6	Responsabilités scientifiques	99
7.7	Liste des publications	100
8	Activités d'enseignement	104
8.1	Enseignements au département informatique de l'IUT de Rodez	105
8.2	Autres enseignements à l'Université Toulouse Capitole	106
8.3	Enseignements à l'Université Toulouse 3 Paul Sabatier	106
8.4	Autres enseignements	107
8.5	Mise en œuvre de pédagogies actives	107
	Bibliographie	109

Table des figures

1.1	Axes, problématiques de recherche et principales contributions.	2
2.1	Représentation BPMN d'un processus d'évaluation formative mobilisant des systèmes de votes interactifs	18
2.2	Représentation BPMN du processus en N Phases	21
2.3	Tsaap-Notes : Fourniture de la première réponse	22
2.4	Tsaap-Notes : confrontation de point de vue et opportunité d'améliorer sa réponse	22
2.5	Tsaap-Notes : histogramme des votes et évaluation par les pairs	23
2.6	Tsaap-Notes : algorithme d'appariement des réponses pour la confrontation de point de vue	24
2.7	Le cycle des <i>Learning Analytics</i> (CLOW, 2012)	32
2.8	Elaastic - La phase 2 composée d'une interface pour l'évaluation par les pairs de réponses correctes et incorrectes.	33
2.9	Échantillon 2019 - Séquences et lignes de la taille de l'effet de Cohen d : aucun effet ($d = 0$), petit ($d = 0,2$), moyen ($d = 0,5$), grand ($d = 0,8$) et très grand ($d = 1.2$).	36
2.10	Modèle d'orchestration de Vickrey élaboré pour le <i>Peer Instruction</i>	38
2.11	Nouveau modèle d'orchestration pour processus à 2 votes	39
2.12	Elaastic - Explication de premier niveau	40
2.13	Elaastic - Explication de niveau 2	40
2.14	Elaastic - Explication avec haut niveau de détail	41
2.15	Elaastic - Recommandations sur les explications à discuter en priorité	41
2.16	Jeu de données 2022 - Séquences et lignes de la taille de l'effet de Cohen d : aucun effet ($d = 0$), petit ($d = 0,2$), moyen ($d = 0,5$), grand ($d = 0,8$) et très grand ($d = 1.2$).	43
2.17	Échantillon 2022 - Répartition des séquences comparée à la répartition attendue	44
3.1	<i>Note as feedback</i> : affichage des résultats d'une question avec note prise par un étudiant.	57

TABLE DES FIGURES

3.2	<i>Note as feedback</i> : la fonctionnalité de génération du fichier au format Moodle.	58
3.3	<i>Note as feedback</i> : édition d'une question issue de Tsaap-Notes et importée dans Moodle.	59
3.4	<i>Note as feedback</i> : affichage du feedback dans Moodle pour une réponse donnée par un étudiant participant à un test de révisions.	59
3.5	<i>Rationale as feedback</i> : affichage du feedback dans Moodle pour une réponse donnée par un étudiant participant à un test de révision. Seules les explications ayant une évaluation moyenne supérieure à 2,5/5 sont affichées.	62
3.6	Prototype haute fidélité présenté durant les entretiens : les 4 écrans principaux	69
4.1	Représentation de la situation prototypique proposée par TCHOUNIKINE .	74
4.2	Auteurs et concepts positionnés suivant 2 dimensions : la discipline dominante (de la psychologie aux neurosciences) et le domaine d'étude dominant (des émotions à la cognition)	76
4.3	Représentation de la nouvelle situation prototypique	78
5.1	Nouveaux axes et nouvelles problématiques de recherche en IEIAH.	84

Liste des tableaux

2.1	Grille de critères d'évaluation des dispositifs étudiés	19
2.2	Évaluation de la plateforme Tsaap-Notes	27
2.3	Comparaison entre les phases dans Tsaap-Notes et Elaastic	34
3.1	Profil des enseignants participant à l'étude	66
3.2	Questionnaire pour les enseignants	67
3.3	Profil des élèves participant à l'étude	67
3.4	Questionnaire pour les élèves	68
3.5	Hypothèses identifiées à l'issue des entretiens	68
3.6	Impact des hypothèses identifiées lors des entretiens sur la conception de l'application Konsolidation	70
4.1	Liens entre interactions Elaastic et états cognitifs/émotionnels impactés.	78

Chapitre 1

Introduction

Contenu

1.1	Contexte général	1
1.2	Axes de recherche et problématiques	2
1.2.1	La technologie en soutien de l'évaluation formative en classe . .	3
1.2.2	La technologie en soutien de l'évaluation formative hors classe	4
1.3	Méthode de conduite de la recherche	4
1.3.1	Posture épistémologique	4
1.3.2	Méthode	6
1.3.3	Éthique de l'observation instrumentée	6
1.4	Organisation du manuscrit	7

1.1 Contexte général

Ce document présente une synthèse des travaux de recherche auxquels j'ai participé activement depuis le début de ma thèse en 2012 jusqu'à la rédaction de ce manuscrit en 2023.

J'ai réalisé ma thèse à l'Université Toulouse 3 Paul Sabatier au sein de l'équipe SIERA de l'Institut de Recherche en Informatique de Toulouse (IRIT, Unité Mixte de Recherche 5505). J'étais alors recruté par l'Université à mi-temps sur un poste d'enseignant-chercheur associé. De septembre 2015 à août 2018, j'ai été recruté à temps plein par la même Université, toujours sur un contrat à durée déterminée. Mes travaux de recherche se sont poursuivis dans le même contexte. J'ai été recruté en septembre 2018 comme Maitre de conférences à l'IUT de Rodez, composante de l'Université Toulouse Capitole. Je suis resté membre de l'IRIT en rejoignant, à sa création, l'équipe TALENT dirigée par Julien Broisin.

Mes recherches s'inscrivent dans le champ de l'Ingénierie des Environnements Informatiques pour l'Apprentissage Humain (IEIAH). La recherche en IEIAH « renvoie aux

travaux dont l'objet est d'étudier les questions scientifiques liées aux concepts, méthodes, théories, techniques et technologies utiles à la conception des EIAH et des logiciels supports à la gestion des situations pédagogiques » (TCHOUNIKINE, 2009, p. 23).

1.2 Axes de recherche et problématiques

TCHOUNIKINE (2009, p. 21) précise que la situation prototypique d'une recherche en IEIAH est la conception d'un logiciel pour une situation pédagogique cible. Cette approche a prévalu dans le cadre des travaux présentés dans ce document, et cela dès ma thèse débutée en 2012. Les 2 axes de recherche structurant cette section correspondent donc à des travaux de recherche portant sur la conception de systèmes technologiques soutenant 2 situations pédagogiques spécifiques. Ces 2 situations sont caractérisées par la mobilisation de processus d'évaluation formative.

La figure 1.1 illustre de manière synthétique comment s'articulent les axes et les problématiques de recherche exposées dans ce manuscrit. Elle présente aussi les principales contributions et des indications temporelles permettant de lier les travaux à leur contexte.

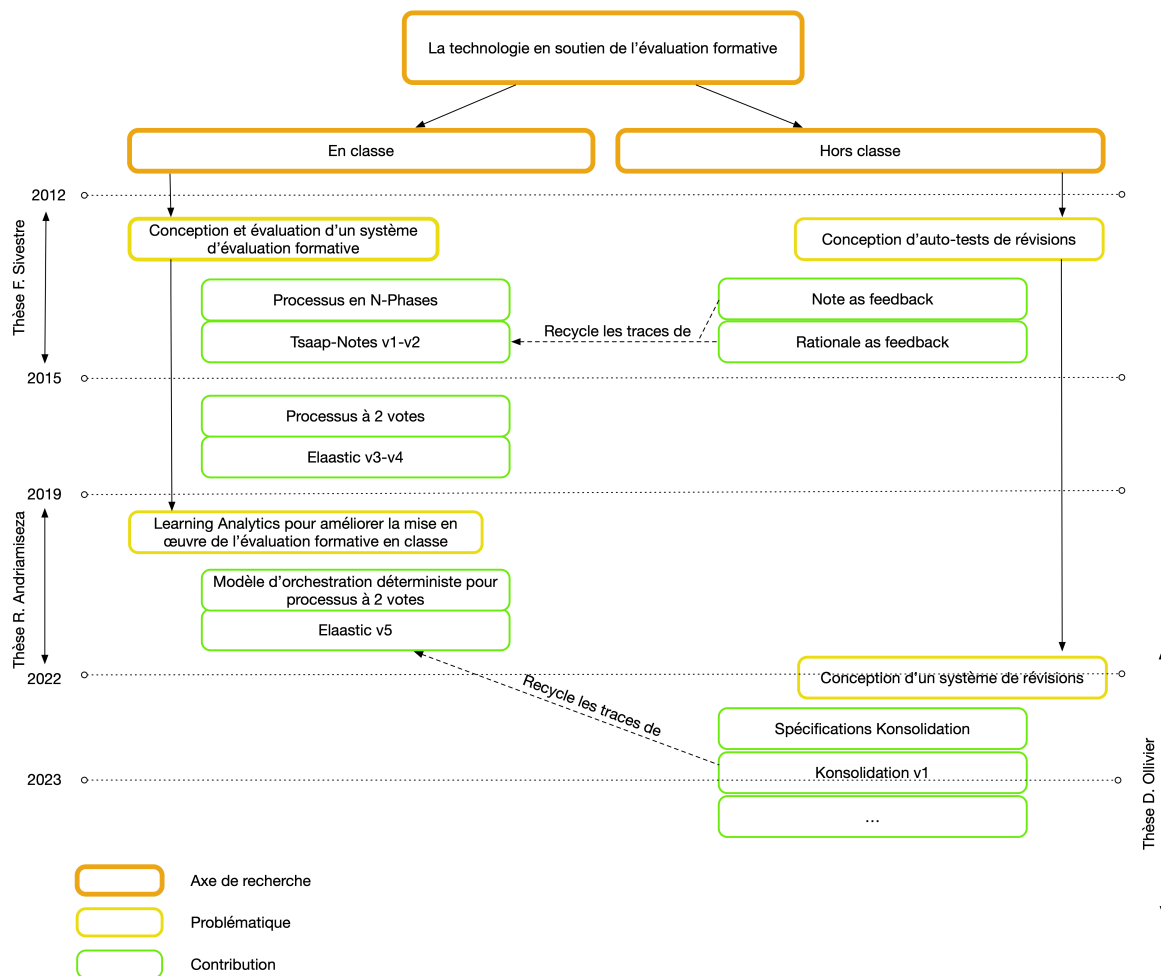


FIGURE 1.1 – Axes, problématiques de recherche et principales contributions.

1.2.1 La technologie en soutien de l'évaluation formative en classe

L'objet principal du travail mené durant ma thèse était de concevoir un système interactif pour faciliter la mise en œuvre de séquences d'évaluation formative dans des cours dispensés en face à face dans le supérieur (SILVESTRE, 2015). En prenant appui sur la théorie de l'évaluation formative formalisée par BLACK et WILIAM (2009), il s'agissait alors, dans cette situation pédagogique cible, de dépasser les limites des systèmes de types *clikers* en enrichissant les interactions humain-machine, les interactions humain-humain et les feedbacks proposés aux étudiants. Les contributions principales de ces travaux sont (1) la formalisation du processus en N Phases permettant de couvrir l'essentiel des activités et stratégies caractéristiques de l'évaluation formative et (2) le système informatique Tsaap-Notes soutenant ce processus. La plateforme Tsaap-Notes a servi de socle au développement d'Elaastic (SILVESTRE et al., 2012), le système technologique sur lequel s'appuie toutes nos expérimentations en lien avec l'évaluation formative depuis 2015.

Ces travaux se sont poursuivis dans le cadre de la thèse de Rialy Andriamiseza co-dirigée avec Julien Broisin et menée en partenariat avec l'Académie de Nancy-Metz. La thèse a débuté en 2019 et a été soutenue en novembre 2022 (ANDRIAMISEZA, 2022). Les travaux ont consisté à analyser les données d'apprentissages (*learning analytics*) collectées avec la plateforme Elaastic entre 2015 et 2019 dans le but d'améliorer la compréhension des séquences d'évaluation formative s'inscrivant dans un processus à 2 votes. Dans un tel processus, les apprenants répondent 2 fois à la même question : une première fois après un temps de réflexion individuelle, et une seconde fois après une confrontation de points de vue entre pairs pouvant prendre la forme d'un micro-débat comme c'est le cas dans le *Peer Instruction* (MAZUR, 1997) ou sous la forme d'une évaluation par les pairs comme cela est le cas dans Elaastic. Ce travail d'analyse exploratoire a permis d'identifier un ensemble d'indicateurs liés à la qualité d'une séquence. Un ensemble de recommandations aux enseignants et concepteurs de systèmes a été formulé pour proposer un modèle d'orchestration complétant les propositions antérieures décrites dans la littérature scientifique. Le modèle, implanté sous forme de recommandations adressées aux enseignants lors du déroulé de la séquence, a été déployé auprès d'enseignants de collège. Les données collectées durant l'expérimentation ont permis de confirmer la validité des indicateurs pour une utilisation en collège, mais n'ont pas permis de mettre à l'épreuve l'efficacité du modèle en raison de la non-utilisation des recommandations dans Elaastic par les enseignants.

1.2.2 La technologie en soutien de l'évaluation formative hors classe

Durant ma thèse, j'ai travaillé à la manière de prolonger les activités formatives hors classe en recyclant les contenus générés durant les séquences d'évaluation formative orchestrées en face à face. L'idée principale étudiée a été le recyclage des questions pour composer des tests de révision accompagnés du recyclage des meilleures explications fournies par les étudiants pour constituer des feedbacks textuels dans ces mêmes tests (SILVESTRE et al., 2015b).

Ces travaux initiant notre engagement dans la problématique de conception de systèmes de révisions se poursuit actuellement dans la cadre de la thèse CIFRE de Denis Ollivier codirigée avec Julien Broisin, co-encadrée par Jean-Baptiste Raclet et menée en partenariat avec l'entreprise Kosmos¹. Le travail de Denis porte non plus uniquement sur la manière de générer des tests de révisions, mais sur la conception d'un système interactif complet soutenant les révisions des apprenants dans un contexte scolaire. Le caractère bénéfique sur les apprentissages des effets de test et d'espacement (LATIMIER, 2019) nous ont conduits à diriger la conception d'un tel système en intégrant le soutien de ces effets et à en étudier l'impact en contexte écologique (OLLIVIER et al., 2023a, 2023b).

1.3 Méthode de conduite de la recherche

1.3.1 Posture épistémologique

Mes formations scientifiques et philosophiques m'ont initié aux pensées des philosophes et historiens des sciences Karl Popper, Thomas Kuhn, et Imre Lakatos. Même si d'autres auteurs tels que Herbert Simon ou Jean-Louis Le Moigne ont permis d'éclairer les spécificités des sciences artificielles dont relève la recherche en IEIAH, ce sont les travaux ancrés dans le post-positivisme qui influencent profondément la manière de conduire ma recherche aujourd'hui.

Je crois à l'existence d'une réalité extérieure au sujet. Je crois à la complexité de la réalité et au besoin de modèles théoriques rendant intelligible le monde de telle sorte que nous puissions y découvrir de nouvelles manières d'agir en son sein. Je crois que la seule manière d'obtenir des modèles de plus en plus pertinents pour résoudre nos problèmes est de les confronter à la réalité dans le contexte d'expérimentations visant à réfuter ces modèles plutôt qu'à les vérifier comme le stipule POPPER (1973).

Le critère de réfutabilité pour distinguer une démarche scientifique des démarches pseudo scientifiques est très discuté : cette approche est adaptée aux sciences expérimentales, c'est-à-dire aux sciences qui peuvent créer des conditions de mise à l'épreuve

1. <https://www.kosmos-education.com/>

des hypothèses formulées dans des conditions parfaitement contrôlées. Toutes les sciences ne disposent pas de cette possibilité, comme l'ethnologie pour n'en citer qu'une. Même dans des champs des sciences les plus « dures » comme en astrophysique, certains modèles ne sont pas testables en environnement contrôlé : la théorie des multivers ne pourra probablement jamais faire l'objet d'une expérience falsificatrice.

Les travaux de KUHN (2008) ont bien mis en lumière le soupçon d'angélisme présent dans la théorie de Popper. Quelle que soit la science, expérimentale ou non, celle-ci est faite par des humains soumis à des contraintes psychologiques et sociologiques qui peuvent rentrer en conflit avec le principe au cœur de la théorie de Popper. Kuhn, parmi toutes les anecdotes fourmillant dans son livre, rappelle la sortie en 1931 du livre « 100 scientifiques contre Einstein » après que ce dernier ait publié en 1905 et 1915 les théories de la relativité restreinte et générale.

Il faudra attendre LAKATOS (1994) pour obtenir une synthèse des postures de Popper et de Kuhn. Sa proposition de structuration de la recherche autour de programmes, composés d'un noyau théorique dur et d'une couche périphérique pouvant « encaisser » les expériences renforçant ou limitant la théorie, offre un cadre permettant de faire cohabiter les pensées de Popper et de Kuhn.

En sciences de l'éducation et dans la communauté EIAH, le critère de démarcation entre sciences et pseudo-sciences proposé par POPPER et la primauté induite sur les expérimentations contrôlées reste une question vive : en 2020, suite aux déclarations du Conseil Scientifique de l'Éducation Nationale indiquant que « seule l'expérimentation contrôlée permet de vérifier qu'un outil pédagogique fonctionne » (CSEN, 2020, p. 12), l'ATIEF revendiqua dans un avis de son conseil d'administration que l'éducation, en tant que fait « social total », ne pouvait se contenter des disciplines s'appuyant sur des expérimentations contrôlées pour comprendre et expliquer les phénomènes observés (ATIEF, 2020).

Je souscris aux critiques du modèle de Popper. Cependant, je ne peux me détacher du caractère efficace de ce modèle pour m'éloigner, en tant que chercheur, de la tentation d'une approche se disant scientifique, mais ne prenant pas toutes les mesures permettant de limiter les impacts des contraintes psychologiques et sociologiques sur la science en train de se faire. Cet attachement au critère de Popper, tout en acceptant ses limites, permet de situer ma posture épistémologique dans un certain périmètre : un post-positivisme non dogmatique dans lequel la pensée de Popper est d'abord un repère éthique.

Cette posture se rapproche très fortement du réalisme critique tel que décrit par CRESWELL et CLARK (2017, p. 42) : le réalisme critique intègre une ontologie réaliste (il existe un monde réel qui existe indépendamment de nos perceptions, théories et constructions) avec une épistémologie constructiviste (notre compréhension du monde est inévitablement une construction bâtie à partir de nos propres perspectives et points de vue).

1.3.2 Méthode

Notre travail de recherche, dès la thèse, s'est inscrit dans une démarche itérative et incrémentale très influencée par notre expérience dans le développement logiciel et plus spécifiquement dans la mise en œuvre des méthodes agiles et du *Lean Startup* (RIES, 2012).

L'évolution vers une approche de recherche fondée sur la conception (DBR pour *Design-based research*) (ANDERSON & SHATTUCK, 2012) s'est faite de manière progressive au fil des encadrements de thèse. Le DBR nous est apparu comme la transposition pertinente des principes d'ingénierie proposés dans les approches de conception centrée humain (HCD pour *Human Centered Design*) au contexte de la recherche.

La mobilisation de méthodes mixtes, articulant des études qualitatives et quantitatives selon des plans séquentiels exploratoires ou explicatifs (CRESWELL & CLARK, 2017), nous ont aidé progressivement à cadrer l'implication des parties prenantes dans les différentes activités d'un cycle DBR. L'alternance entre études qualitatives et quantitatives incarne aussi cette posture post-positiviste (importance des études quantitatives fondées sur des données objectives) et non dogmatique (importance des études qualitatives pour faciliter l'émergence d'explications des phénomènes complexes étudiés).

1.3.3 Éthique de l'observation instrumentée

Nos travaux de recherche s'appuient essentiellement sur la collecte de traces. Nous sommes amenés à collecter et à traiter des données à caractère personnel. Quand cette collecte est instrumentée avec des systèmes interactifs, et ainsi automatisée, diverses questions d'ordre éthique se posent. Ces questions sont, pour la plupart, en lien avec le traitement des données à caractère personnel.

Sur l'ensemble de la période couvrant les travaux de recherche auxquels j'ai participé, j'ai fait évoluer la méthode de conduite de ma recherche sur les deux questions suivantes :

- S'est-on assurés que les personnes observées savent réellement ce qui est collecté, par qui ou quoi, pour le compte de qui, quand et dans quel objectif ?
- S'est-on assurés que les personnes observées (enseignants, apprenants et responsables légaux des apprenants si besoin) ont consenti de manière éclairée à la collecte des données ?

Trois étapes clés jalonnent cette évolution. Initialement, une déclaration rigoureuse auprès de la Commission Nationale de l'Informatique et des Libertés (CNIL) était notre pratique standard. Avec l'adoption au niveau européen du Règlement Général sur la Protection des Données (RGPD), nous avons adapté nos systèmes interactifs pour être en pleine conformité avec cette nouvelle réglementation. Enfin, nous avons intégré l'approche *Privacy by Design* dans la conception de nouveaux systèmes interactifs, mettant l'accent sur la protection de la vie privée dès les premières étapes de développement.

Cette démarche d'amélioration continue a renforcé notre collaboration avec le service de notre laboratoire spécialisé dans ces domaines, assurant ainsi une vigilance constante sur les évolutions légales et les questions éthiques.

1.4 Organisation du manuscrit

Le manuscrit est organisé en trois parties.

La première partie présente en détail les travaux de recherche menés dans le champ de l'évaluation formative soutenue par la technologie. Cette partie est structurée en deux chapitres (2 et 3). Ils permettent de distinguer les travaux en lien avec l'évaluation formative orchestrée par l'enseignant pendant ses cours des travaux portant sur l'évaluation formative se déroulant hors classe dans le contexte d'activités de révisions. Ces travaux ont été initiés durant ma thèse codirigée par Philippe Vidal et Julien Broisin. Ils se sont poursuivis jusqu'à aujourd'hui dans le cadre de thèses codirigées avec Julien Broisin. Ils sont le socle du projet de recherche présenté dans la deuxième partie.

Cette deuxième partie du document est composée de 2 chapitres. Le premier chapitre propose une relecture des travaux antérieurs afin d'introduire et de justifier une nouvelle situation prototypique de la recherche en IEIAH. Le deuxième chapitre présente alors le projet de recherche enclenché depuis quelques mois et permettant d'alimenter mon activité sur les 5 à 10 ans à venir. Ce projet de recherche, décrit sous l'angle de la nouvelle situation prototypique d'une recherche en IEIAH, s'intéresse à la manière dont la technologie peut soutenir les états psychologiques et émotionnels des apprenants favorables à l'apprentissage tout en limitant ceux représentant des freins à l'apprentissage. Deux axes principaux sont exposés pour mener à bien ces recherches : le premier mobilise les pairs comme ressource d'instruction, le second vise à explorer les bénéfices éventuels apportés par les intelligences artificielles génératives dans la poursuite de notre objectif.

La troisième et dernière partie présente le rapport de mes activités depuis ma thèse jusqu'à la rédaction de ce document. Le rapport porte sur les activités de recherche et d'enseignement. Le *curriculum vitae* résumé, fourni en introduction de cette partie, présente une synthèse de toute ma carrière professionnelle incluant ainsi un aperçu de mes activités en dehors du monde académique.

Première partie
Travaux de recherche

Chapitre 2

La technologie en soutien de l'évaluation formative en classe

Contenu

2.1	État de l'art	10
2.1.1	Les notions clés	10
2.1.2	Les évaluations formatives et leur mise en œuvre dans la classe	11
2.1.3	Les évaluations formatives soutenues par la technologie	12
2.2	Problématiques et questions de recherche	14
2.2.1	Conception d'un système d'évaluation d'audience	14
2.2.2	Compréhension et amélioration des processus à 2 votes	16
2.3	Conception d'un système d'évaluation d'audience	17
2.3.1	Méthode de conduite de la recherche	17
2.3.2	Identification des processus d'évaluation d'audience	17
2.3.3	Évaluation et comparaison des processus d'évaluation d'audience	18
2.3.4	Le processus en N Phases	20
2.3.5	Tsaap-Notes	20
2.3.6	Expérimentations et résultats	24
2.3.7	Discussion et positionnement des travaux	26
2.3.8	Rétrospective	29
2.4	Compréhension et amélioration des processus à 2 votes	31
2.4.1	Méthode de conduite de la recherche	31
2.4.2	L'analyse de données exploratoire	32
2.4.3	Résultats et recommandations induites	35
2.4.4	Un nouveau processus à 2 votes	38
2.4.5	Expérimentation et résultats	41
2.4.6	Discussion et positionnement des travaux	44
2.4.7	Rétrospective	47

2.1 État de l'art

2.1.1 Les notions clés

En 2007, WILIAM introduit les expressions « évaluation de l'apprentissage » et « évaluation pour l'apprentissage » pour distinguer respectivement évaluation sommative et évaluation formative (WILIAM, 2007). Les évaluations sommatives ont pour objectif la mesure du niveau d'atteinte par les étudiants des objectifs d'un programme scolaire ; elles permettent, à la fin d'une unité d'enseignement, de comparer le niveau ou la performance de l'étudiant par rapport à un niveau de référence attendu en vue de l'éventuelle délivrance d'une certification (SADLER, 1989). L'évaluation formative, quant à elle, s'inscrit dans une démarche de régulation des apprentissages (PERRENOUD, 1998 ; WILIAM, 2007) poursuivant ainsi un autre objectif décrit dans (CARNEGIE MELLON UNIVERSITY, s. d.) :

« L'objectif des évaluations formatives est de mesurer le niveau d'apprentissage des étudiants pour fournir un feedback qui peut être utilisé par les enseignants pour améliorer leur enseignement et par les étudiants pour améliorer leur apprentissage. De manière plus spécifique, les évaluations formatives :

- aident les étudiants à identifier leur force et leur faiblesse et à cibler les sujets qui nécessitent plus de travail ;
- aident les enseignants à reconnaître les difficultés des étudiants et à traiter les problèmes immédiatement. »

Cette définition explicite le rôle clé joué par le concept de feedback au cœur de l'évaluation formative, et corrobore ainsi l'idée que les concepts de feedback et d'évaluation formative sont étroitement liés (BLACK & WILIAM, 1998).

Selon SADLER, un feedback est une information portant sur l'écart entre un niveau actuel et un niveau de référence. Dans le contexte d'une évaluation formative, le niveau observé est le niveau d'atteinte par un apprenant d'un ou de plusieurs objectifs d'un enseignement ; cette information est utilisée pour modifier cet écart d'une quelconque manière (SADLER, 1989).

SADLER identifie trois conditions nécessaires pour que les apprenants tirent un bénéfice d'un feedback dans le contexte d'une tâche d'apprentissage ; il déclare qu'un apprenant doit connaître :

- ce qu'est la performance attendue ;
- comment se positionne sa performance actuelle par rapport à la performance attendue ;

- comment agir pour réduire l'écart entre sa performance actuelle et la performance attendue.

Par ailleurs, les études sur l'engagement cognitif des apprenants montrent que l'argumentation entre pairs, plus généralement la possibilité de s'appuyer sur les connaissances des autres apprenants pour construire sa propre connaissance conduit à un engagement cognitif important (CHI et al., 2018) et donc à de meilleurs résultats d'apprentissage. En plus du feedback « traditionnel », l'introduction de phases d'interactions entre pairs au sein des évaluations formatives, quelle que soit la forme, orale ou écrite, que prennent ces interactions, est une clé de succès dans leurs mises en œuvre.

L'évaluation formative et la fourniture de feedback *ad hoc* sont des leviers pour améliorer les résultats d'apprentissage, quels que soient le niveau et les disciplines dans lesquels elles opèrent (BLACK & WILIAM, 2009 ; HATTIE & TIMPERLEY, 2007). L'évaluation formative est donc idéalement mise en œuvre de manière fréquente et systématique.

2.1.2 Les évaluations formatives et leur mise en œuvre dans la classe

BLACK et WILIAM (2009), dans le cadre de leur théorie de l'évaluation formative, ont identifié 5 activités de référence dans la mise en œuvre d'évaluations formatives :

- le partage des critères de succès avec les apprenants ;
- le questionnement à la classe ;
- l'évaluation sous forme de commentaires plutôt qu'à l'aide de notes chiffrées uniquement ;
- l'auto-évaluation et l'évaluation par les pairs ;
- l'utilisation de tests sommatifs en vue d'usage formatif.

Ces 5 activités s'inscrivent plus largement dans 5 stratégies de mise en œuvre que BLACK et WILIAM formulent comme suit :

1. clarifier et partager les objectifs d'apprentissage et les critères de succès ;
2. concevoir des situations efficaces de discussions en classe et d'autres tâches d'apprentissage qui explicitent les preuves du niveau d'avancement des apprenants dans leur apprentissage ;
3. fournir des feedbacks qui permettent aux apprenants de progresser dans leur apprentissage ;
4. faire en sorte que les apprenants deviennent des ressources d'instruction pour les autres apprenants ;
5. faire en sorte que les apprenants deviennent responsables de leur propre apprentissage.

Concrètement, l'évaluation formative peut se décliner à l'oral : en face à face, c'est ce qui se produit quand un enseignant interroge un ou plusieurs élèves en classe. La réponse fournie par un élève, censée refléter son niveau sur l'objectif d'apprentissage ciblé, est analysée et commentée par les pairs ou directement par l'enseignant. Ces échanges et l'intervention de l'enseignant constituent le feedback adressé aux élèves. Dans ce contexte, plus le nombre d'élèves est important, plus il est difficile de collecter et de traiter l'ensemble des réponses apportées par les élèves. Pour une question donnée, il est alors très probable que seul un petit lot d'élèves soit engagé fortement dans la séquence d'évaluation formative.

Une autre forme de pratique consiste à donner un travail écrit à réaliser en classe ou à la maison qui sera relevé et corrigé par l'enseignant afin de fournir un feedback personnalisé. Cette forme d'activité peut être dispensée entièrement à distance ou dans un contexte hybride : les élèves travaillent en dehors de la classe et tout ou partie de la restitution s'effectue en face à face.

Cette démarche vise l'engagement de tous les élèves dans des activités d'apprentissage, mais pose au moins deux problèmes :

- le feedback est fourni loin du temps de l'activité, ce qui peut nuire à son efficacité (HATTIE & TIMPERLEY, 2007) ;
- le travail de l'enseignant est difficilement soutenable si l'objectif est une mise en œuvre fréquente et systématique de l'évaluation formative.

En synthèse, la mise en œuvre de l'évaluation formative augmente en complexité dès lors que le nombre d'élèves augmente dans la classe. Des systèmes technologiques sont apparus pour collecter et analyser un grand nombre de réponses obtenues à une question posée à la classe afin de fournir, au bon moment, un feedback adapté, c'est-à-dire aidant, *in fine*, l'apprenant à apprendre.

2.1.3 Les évaluations formatives soutenues par la technologie

Les systèmes de votes interactifs

Les systèmes de votes interactifs permettent à une classe entière ou un amphithéâtre de répondre à une question à choix multiples affichée sur un écran (KAY & LESAGE, 2009). Chaque étudiant utilise un dispositif (boîtier de vote, plus récemment smartphone, tablette ou ordinateur) lui permettant de saisir la réponse à la question. Les réponses sont collectées en temps réel et les résultats affichés après que l'ensemble des étudiants ont répondu à la question. Les résultats sont en général présentés sous forme d'histogrammes et préservent l'anonymat des participants. Diverses études ont montré les bénéfices apportés par l'utilisation des systèmes de votes interactifs. Ceux-ci s'avèrent très efficaces pour motiver et engager les étudiants (GAUCI et al., 2009 ; UHARI et al., 2003) ainsi que

pour améliorer les résultats des étudiants participant aux votes interactifs (SHAFFER & COLLURA, 2009). L'étude de GAUCI et al. insiste sur le fait que les résultats obtenus permettent d'orienter les discussions pendant le cours. Ces résultats donnent l'occasion à tous les participants (enseignants et apprenants) de donner et de recevoir plus d'explications ou de résoudre des conflits sociocognitifs.

L'instruction par les pairs

L'approche baptisée « instruction par les pairs » (IP – « *Peer Instruction* » en anglais) a été introduite par MAZUR pour la première fois en 1991 durant des cours de physique pour des étudiants de licence. L'approche consiste, pour une même question de compréhension (issue d'un *conceptest*), à sonder les étudiants deux fois : une première fois après une phase de réflexion individuelle, puis une deuxième fois après que les étudiants ont débattu avec leurs voisins du bien-fondé de la réponse qu'ils ont fournie à l'issue du premier sondage. À la suite du deuxième vote, un feedback collectif est donné par l'enseignant sur la réponse correcte et le raisonnement attendu. Les résultats du deuxième sondage révèlent d'une part un plus grand nombre de bonnes réponses, mais surtout que les bénéfices en termes d'apprentissage se trouvent renforcés par rapport à une approche classique (MAZUR, 1997).

Devant les résultats probants apportés par l'approche IP (CROUCH & MAZUR, 2001 ; SMITH et al., 2011), beaucoup d'universités se sont lancées dans l'utilisation des systèmes de votes interactifs en appliquant la méthode de MAZUR (BORREGO et al., 2011 ; BURNSTEIN & LEDERMAN, 2001 ; CLINE, 2006 ; FAGEN et al., 2002). En France, deux universités ont conduit des études sur le PI (PARMENTIER et al., 2015 ; RUDOLPH et al., 2014).

En forçant la discussion entre pairs entre les deux phases d'interrogation de l'assemblée, l'approche IP renforce d'une part la mise en place de situations de discussions en classe, et d'autre part le fait que les étudiants deviennent des ressources d'instruction pour les autres.

Le développement des services Web

Le développement des technologies permet la mise en œuvre d'évaluations formatives à grande échelle. Le développement d'Internet a permis de porter les dispositifs technologiques initiaux sur le Web rendant accessibles les services offerts *via* tout dispositif connecté au réseau des réseaux : ordinateur, tablette, smartphone.

Divers outils permettent la mise en place de tests autocorrectifs. Les principales plateformes de cours en ligne telles que Moodle mettent à disposition des services de création et de passage de tests autocorrectifs. La plupart des environnements numériques de travail fournissent aussi de tels dispositifs. Enfin, le réseau Canopé propose le service en ligne

Quizinière¹.

Les services Kahoot², Socrative³, Wooclap⁴ et Plickers⁵ sont des exemples (liste non exhaustive) de services Web soutenant la mise en œuvre d'évaluations formatives sur le modèle des systèmes de votes interactifs. Les études portant sur ces outils pointent des bénéfices similaires à l'utilisation des dispositifs antérieurs : engagement et motivation renforcés pour les étudiants (CHNG & GURVITCH, 2018; WANG & TAHIR, 2020; WASH, 2014).

Cette transition vers les services Web a permis d'une part un enrichissement des interactions et d'autre part une mobilisation de ces technologies dans des contextes d'apprentissage à distance ou hybride.

2.2 Problématiques et questions de recherche

2.2.1 Conception d'un système d'évaluation d'audience

L'institut de statistique de l'UNESCO (2022) indique que le nombre d'étudiants dans les universités du monde entier a doublé entre 2000 et 2020. La baisse de 1,5% constatées en France pour la rentrée 2023 (LE MONDE, 2023), reste négligeable en rapport avec la croissance constante des 20 dernières années à l'échelle mondiale. Cette croissance pose naturellement la question d'un modèle d'enseignement soutenable à l'échelle des états.

L'enseignement en amphithéâtre, permettant à un enseignant de s'adresser simultanément à plusieurs dizaines ou centaines d'étudiants, est séduisant dans un tel contexte. L'introduction des systèmes de votes interactifs dans les amphithéâtres ont permis d'améliorer l'engagement et les résultats des étudiants dans les formations concernées (GAUCI et al., 2009; SHAFFER & COLLURA, 2009). L'Instruction par les pairs (CROUCH & MAZUR, 2001), en introduisant les microdébats sur les questions de compréhension entre 2 votes, conduit à un processus sophistiqué d'orchestration de séquences d'évaluation formative (VICKREY et al., 2015) et engage les étudiants sur le plus haut niveau (Interactif) de l'échelle ICAP (CHI et al., 2018). Les boîtiers de votes et l'instruction par les pairs, en permettant la mise en œuvre soutenable d'évaluations formatives engageantes durant les cours en face à face à très large audience, m'ont incité à poursuivre des recherches relatives à l'enseignement en amphithéâtre.

La problématique de recherche à laquelle je m'intéresse, d'abord durant ma thèse, puis durant la codirection de la thèse de Rialy Andriamiseza, est la conception de systèmes interactifs permettant de soutenir l'évaluation formative durant les cours en face

1. <https://www.quiziniere.com/>

2. <https://kahoot.it/>

3. <https://www.socrative.com/>

4. <https://www.wooclap.com/fr/>

5. <https://get.plickers.com/>

à face à large audience. Nous avons baptisé de tels systèmes des *systèmes d'évaluation d'audience*. Ces travaux s'inscrivent d'abord exclusivement dans l'enseignement supérieur dans le cadre des cours dispensés en amphithéâtre avant de concerner aussi l'enseignement secondaire dans le cadre des cours dispensés dans les classes de collège ou lycée.

Les processus d'évaluation formative s'appuyant sur les boîtiers de votes interactifs ou sur les dispositifs connectés (smartphones, tablettes, ordinateurs) mobilisent des interactions circonscrites à la sélection d'un ou de plusieurs items pour répondre à une question à choix ou, parfois, à la saisie de réponses très courtes sur des questions ouvertes.

Les étudiants ne sont pas invités à formaliser par écrit leur raisonnement, et cela met en lumière deux limites :

- Durant les évaluations sommatives, il est très fréquemment demandé aux étudiants d'expliquer par écrit les résultats qu'ils obtiennent. Les approches courantes mobilisant les boîtiers de votes ne sont donc pas optimales en terme d'alignement constructiviste (BIGGS, 1996) : l'activité proposée durant le cours pour atteindre l'objectif d'apprentissage n'est pas de même nature que celle proposée pour évaluer de manière sommative le niveau d'atteinte de l'objectif d'apprentissage.
- Les feedbacks produits par les systèmes sont les résultats de calculs statistiques rudimentaires donnant lieu par exemple à la représentation sous forme d'histogrammes de la distribution des votes ou à la représentation sous forme de nuage de points des mots ou expressions les plus citées dans les réponses textuelles courtes. L'enseignant, dans le meilleur des cas, ne dispose alors que du pourcentage de réponses correctes pour évaluer le niveau de compréhension des étudiants. L'absence d'information sur la manière dont le choix correct a été sélectionné ne permet pas d'éliminer les faux positifs (apprenants ayant répondu correctement au hasard ou sur la base d'un raisonnement erroné).

La question de recherche découlant de ces constats est la suivante :

[QR 1] Comment concevoir un système interactif soutenant l'évaluation formative durant les cours en face à face à large audience dépassant les limites des systèmes courants ?

Afin de répondre aux premières limites identifiées, nous nous intéressons plus spécifiquement à la question suivante :

[QR 1.1] Comment soutenir l'argumentation écrite dans un processus d'évaluation formative mobilisable durant les cours en face à face à large audience ?

Dans le cadre d'un tel processus d'évaluation formative en amphithéâtre, la collecte des justifications écrites des étudiants soulève une première question relative à l'évaluation de

ces nouvelles contributions. En effet, il n'est pas soutenable pour l'enseignant d'évaluer l'ensemble des contributions écrites dans un temps raisonnable. Par ailleurs, nous avons évoqué précédemment un des bénéfices importants du *Peer Instruction* : il introduit les micro-débats dans la séquence pédagogique permettant ainsi d'engager les étudiants au plus haut niveau cognitif suivant l'échelle ICAP (CHI & WYLIE, 2014). Cette approche micro-débat offre une opportunité de susciter des conflits sociocognitifs dans un contexte favorable au développement des apprentissages (BUCHS et al., 2008). Les processus collectant les explications écrites ne devraient pas introduire de régression sur les bénéfices apportés par le *Peer Instruction*.

Nous proposons de lever ces deux nouveaux verrous en ancrant fortement les nouveaux processus dans une démarche répondant à la 5^e stratégie caractéristique de l'évaluation formative identifiée par BLACK et WILIAM (2009) : la mobilisation des pairs comme ressources d'instruction. Cela nous conduit à formuler une nouvelle question de recherche comme suit :

[QR 1.2] Comment soutenir la mobilisation des pairs comme ressource d'instruction dans un processus d'évaluation formative en vue de fournir un feedback sur les contributions textuelles, et de maintenir les bénéfices des interactions entre pairs observés dans le *Peer Instruction* ?

2.2.2 Compréhension et amélioration des processus à 2 votes

Les processus *Peer Instruction*, en N Phases, puis celui implanté dans la plateforme Elaastic ont une caractéristique commune : ils s'appuient d'une part sur des questions à choix sollicitant un vote des étudiants (sur un ou plusieurs choix) pour répondre à chaque question, et d'autre part proposent aux étudiants de répondre 2 fois à la même question à choix durant la séquence d'évaluation formative. Une interaction entre pairs prenant différentes formes selon les processus est introduite entre les deux votes afin de conduire les apprenants à mieux répondre à la question lors de la deuxième tentative. Afin de faciliter le discours sur l'ensemble de ces processus, nous les avons baptisé *processus à 2 votes*.

Les traces collectées avec la plateforme Elaastic dans l'enseignement supérieur entre 2015 et 2019 vont révéler un résultat remarquable : parmi les séquences d'évaluations formatives étudiées, près d'un quart d'entre elles ne peuvent pas être considérées comme bénéfiques pour les étudiants. Ce résultat conduit à la deuxième question de recherche traitée principalement durant la thèse de Rialy Andriamizeza :

[QR 2] Comment tirer parti de l'analyse des données d'apprentissage (*Learning Analytics*) pour améliorer les résultats des séquences d'évaluations formatives s'inscrivant dans les processus à 2 votes ?

Plus précisément, les questions abordées durant la thèse de Rialy sont les suivantes :

[QR 2.1] Quels sont les indicateurs permettant de prédire le succès d'une séquence d'évaluation formative suivant un processus à 2 votes ?

[QR 2.2] Sur la base de ces indicateurs, quelles recommandations adressées aux enseignants pour orchestrer au mieux une séquence d'évaluation formative suivant un processus à 2 votes ?

[QR 2.3] Sur la base de ces indicateurs, quelles recommandations adressées aux concepteurs de systèmes interactifs pour soutenir au mieux une séquence d'évaluation formative suivant un processus à deux votes ?

2.3 Conception d'un système d'évaluation d'audience

2.3.1 Méthode de conduite de la recherche

Les travaux relatifs à la conception d'un système d'évaluation d'audience ont fait l'objet d'une démarche très exploratoire visant d'abord à répondre une très vaste question : comment engager cognitivement les étudiants durant les cours en face à face dispensés en amphithéâtre ? Les explorations ont conduit progressivement à travailler sur l'évaluation formative qui s'est avérée être le meilleur cadre théorique issu des sciences de l'éducation pour répondre à la question initiale. C'est à la fin de ma thèse que les pièces du puzzle se sont emboîtées pour être en capacité de raconter une histoire satisfaisante dans le contexte d'un manuscrit de thèse. Cette histoire présente une démarche en 5 étapes :

1. L'identification et la représentation des processus d'évaluation d'audience existants.
2. L'évaluation et la comparaison des processus identifiés.
3. La proposition d'un nouveau processus dépassant les limites des processus identifiés.
4. La conception d'un système technologique soutenant la mise en œuvre du nouveau processus.
5. L'expérimentation pour mettre à l'épreuve le nouveau processus et le système l'implantant.

2.3.2 Identification des processus d'évaluation d'audience

Notre approche a consisté dans un premier temps à identifier les processus d'évaluations formatives mobilisés durant les cours en face à face à large audience référencés dans la littérature scientifique et repérés dans une démarche de veille technologique. Ce premier

travail a d'abord conduit à la proposition des définitions d'évaluation d'audience et de système d'évaluation d'audience.

Définition 1 Une évaluation d'audience est un processus d'évaluation formative se déroulant pendant un cours dispensé en face-à-face à un grand nombre d'étudiants et s'adressant simultanément à l'ensemble des étudiants participant au cours.

Définition 2 Un Système d'Évaluation d'Audience (SEA) est un système supportant un processus d'évaluation d'audience.

Notons qu'à partir de cette dernière définition, [QR 1] peut être reformulée de la manière suivante :

[QR 1] Comment concevoir un système d'évaluation d'audience dépassant les limites des systèmes courants ?

Les processus d'évaluation d'audience identifiés ont alors été modélisés à l'aide du formalisme BPMN⁶. La figure 2.1 est une représentation BPMN d'un processus basique d'évaluation formative en amphithéâtre.

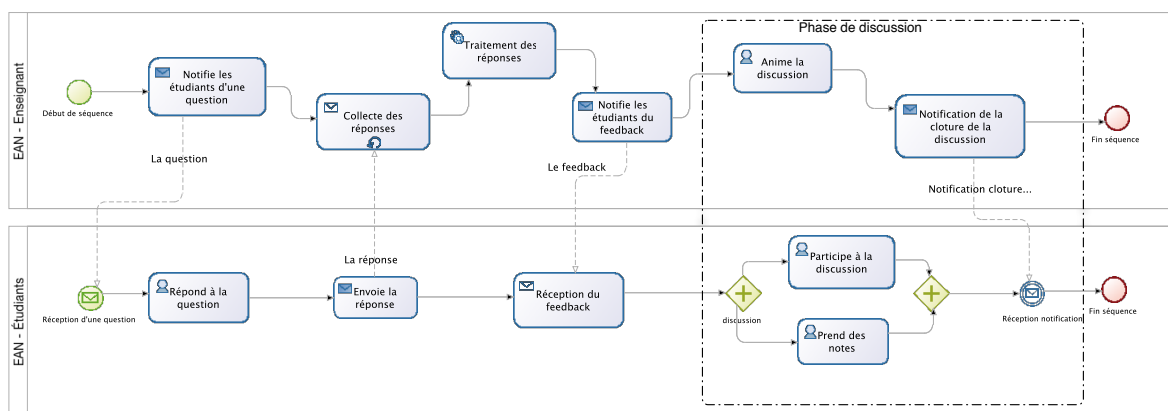


FIGURE 2.1 – Représentation BPMN d'un processus d'évaluation formative mobilisant des systèmes de votes interactifs

2.3.3 Évaluation et comparaison des processus d'évaluation d'audience

Dans un second temps, nous avons proposé un cadre pour évaluer et comparer les processus identifiés. À cette fin, nous nous sommes appuyé sur la théorie de l'évaluation formative formalisée par BLACK et WILIAM (2009). La grille d'évaluation proposée (cf. tableau 2.1) a permis de mettre en lumière les limites des processus identifiés et de guider

6. <https://www.bpmn.org/>

la conception d'un nouveau processus d'évaluation d'audience, baptisé processus en N Phases, et l'implantation de celui-ci dans un système d'évaluation d'audience utilisable en contexte écologique : Tsaap-Notes.

TABLE 2.1: Grille de critères d'évaluation des dispositifs étudiés

Code	Description	Domaine de valeurs						
Activités de référence								
A-1	Le dispositif participe à l'activité de partage des critères de succès avec les apprenants	<table border="1"> <tr> <td></td> <td>ne participe pas</td> </tr> <tr> <td></td> <td>participe</td> </tr> <tr> <td></td> <td>réalise</td> </tr> </table>		ne participe pas		participe		réalise
	ne participe pas							
	participe							
	réalise							
A-2	Le dispositif participe à l'activité de questionnement de la classe							
A-3	Le dispositif participe à l'évaluation sous forme de commentaires plutôt qu'à l'aide de notes chiffrées							
A-4	Le dispositif participe à l'auto-évaluation et l'évaluation par les pairs							
A-5	Le dispositif participe à l'utilisation de tests sommatifs en vue d'usage formatif							
Stratégies								
S-1	Le dispositif participe à la clarification et au partage des objectifs d'apprentissage et des critères de succès	<table border="1"> <tr> <td></td> <td>ne participe pas</td> </tr> <tr> <td></td> <td>participe</td> </tr> <tr> <td></td> <td>réalise</td> </tr> </table>		ne participe pas		participe		réalise
	ne participe pas							
	participe							
	réalise							
S-2.1	Le dispositif participe à la conception de situations efficaces de discussions en classe							
S-2.2	Le dispositif participe à des tâches qui élicitent les preuves du niveau d'avancement des apprenants dans leur apprentissage							
S-3	Le dispositif participe à la fourniture de feedbacks qui permettent aux apprenants de progresser dans leur apprentissage							
S-4	Le dispositif participe au fait que les apprenants deviennent des ressources d'instruction pour les autres apprenants							
S-5	Le dispositif participe au fait que les apprenants deviennent responsables de leur propre apprentissage							
Interactions								
I-ST	Support de la saisie de texte libre par le dispositif	<table border="1"> <tr> <td></td> <td>ne propose pas la saisie de texte</td> </tr> <tr> <td></td> <td>propose la saisie de texte court</td> </tr> <tr> <td></td> <td>propose la saisie de texte libre</td> </tr> </table>		ne propose pas la saisie de texte		propose la saisie de texte court		propose la saisie de texte libre
	ne propose pas la saisie de texte							
	propose la saisie de texte court							
	propose la saisie de texte libre							
I-TF	Types de feedback fournis par le dispositif en sortie	<table border="1"> <tr> <td></td> <td>pas de feedback</td> </tr> <tr> <td></td> <td>le feedback est fourni sous forme d'indicateurs numériques</td> </tr> <tr> <td></td> <td>le feedback est fourni sous forme textuelle</td> </tr> </table>		pas de feedback		le feedback est fourni sous forme d'indicateurs numériques		le feedback est fourni sous forme textuelle
	pas de feedback							
	le feedback est fourni sous forme d'indicateurs numériques							
	le feedback est fourni sous forme textuelle							
Engagement								
E-TE	Le dispositif engage fortement les étudiants	<table border="1"> <tr> <td></td> <td>engage moins d'un tiers des étudiants</td> </tr> <tr> <td></td> <td>engage entre un tiers et deux tiers des étudiants</td> </tr> <tr> <td></td> <td>engage plus de deux tiers des étudiants</td> </tr> </table>		engage moins d'un tiers des étudiants		engage entre un tiers et deux tiers des étudiants		engage plus de deux tiers des étudiants
	engage moins d'un tiers des étudiants							
	engage entre un tiers et deux tiers des étudiants							
	engage plus de deux tiers des étudiants							

2.3.4 Le processus en N Phases

La figure 2.2 fournit une présentation détaillée du processus en N Phases en BPMN. Le processus est composé de 4 phases. La première phase permet aux étudiants de répondre à une question à choix posée par l'enseignant en fournissant une justification écrite et un degré de confiance sur la réponse fournie.

La deuxième phase est la confrontation de point de vue : chaque étudiant accède à une réponse différente de la sienne fournie par un étudiant ayant participé à la même séquence.

La troisième phase expose la répartition des votes à l'issue de la deuxième phase et enrôle les étudiants dans une évaluation par les pairs portant sur les réponses correctes uniquement. L'approche proposée, en plus de mobiliser les capacités métacognitives des étudiants, permet de fournir un dispositif d'évaluation des contributions écrites soutenable à l'échelle d'une séquence pédagogique de quelques minutes concernant un grand nombre d'étudiants. L'enseignant n'étant pas en capacité d'évaluer en temps réel un grand nombre d'explications écrites, le parti-pris a été de mobiliser l'ensemble des étudiants pour rendre cette tâche réalisable dans un délai raisonnable.

La quatrième phase est une phase de discussion animée par l'enseignant. Elle n'est pas représentée sur la figure 2.2 car elle est identique à la phase de discussion représentée sur la figure 2.1.

2.3.5 Tsaap-Notes

Tsaap-Notes⁷ est une application Web *Open Source* développée et déployée dans l'environnement technologique Java EE. Tsaap-Notes permet à l'enseignant d'orchestrer une séquence d'évaluation formative suivant le processus à N Phases et permet aux étudiants de prendre part à chacune des phases du processus.

Phase 1 La figure 2.3 présente l'interface de Tsaap-Notes permettant aux étudiants de répondre à la question posée par l'enseignant en fournissant une justification écrite.

Phase 2 La figure 2.4 présente l'interface de Tsaap-Notes permettant aux étudiants de consulter la réponse alternative et de modifier éventuellement leur réponse. Les réponses alternatives sont proposées de manière anonyme afin de limiter les biais d'influence relationnelle.

Phase 3 La figure 2.5 présente l'interface de Tsaap-Notes permettant l'évaluation par les pairs après restitution de la répartition des votes à la deuxième soumission de réponse.

7. <https://github.com/TSaaP/tsaap-notes>

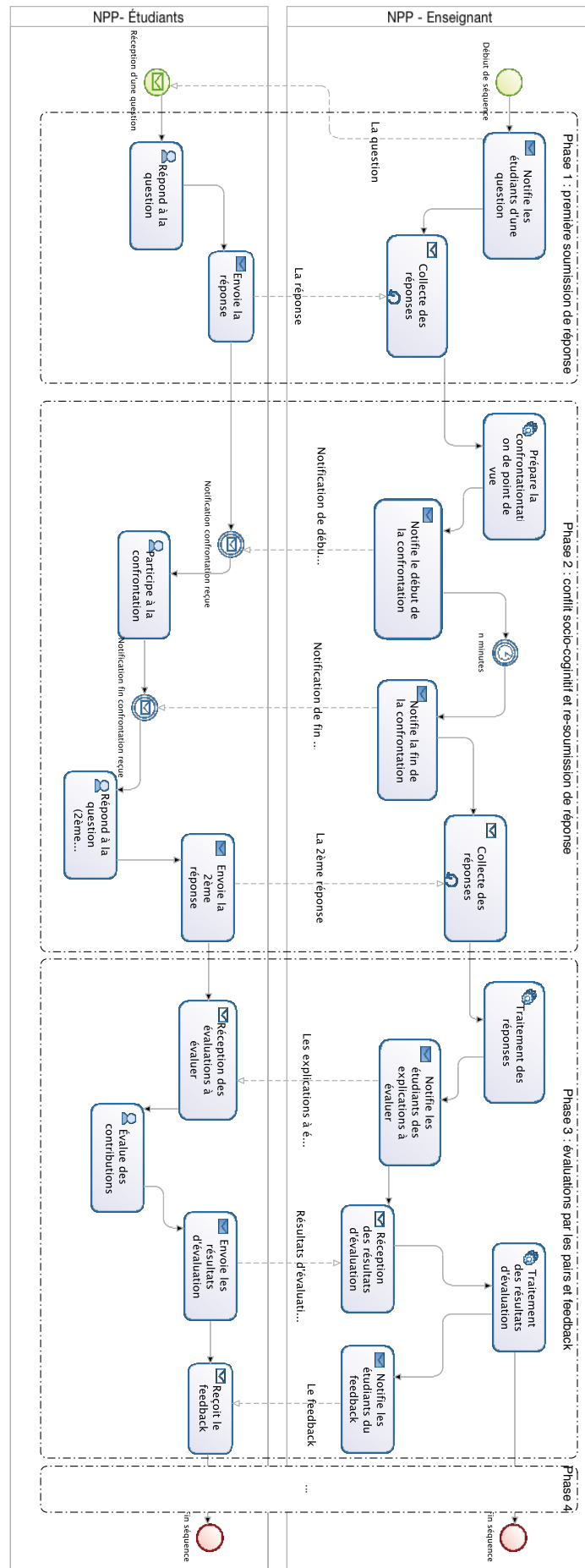


FIGURE 2.2 – Représentation BPMN du processus en N Phases

@FranckSilvestre Master/Informatique/DL/2013_2015/DCLL/1 13 janvier 2015 15:32

La licence GPL et Javascript

Soit L une librairie javascript distribuée sous licence GPL . Soit E une entreprise qui modifie la librairie L pour en faire une version L2. L'entreprise E propose à des utilisateurs son application serveur qui utilise la version L2 de la librairie. Est ce que les utilisateurs peuvent exiger de récupérer le code source de la version L2 du logiciel ?

Oui
 Non

Give an explanation to your choice

What is the degree of confidence in your response (1: not confident to 5:very confident) ? 1

Submit

Reply Favorite I learn

FIGURE 2.3 – Tsaap-Notes : Fourniture de la première réponse

@FranckSilvestre Master/Informatique/DL/2013_2015/DCLL/1 13 janvier 2015 15:32

Here is an alternative response to yours.
Please examine this response and then feel free to change your response, explanation or confidence degree if necessary.

Oui
Le javascript étant exécuté côté client, l'utilisateur a une copie, la contamination du GPL force donc la possibilité d'accès au code à l'utilisateur.

Réponse alternative

La licence GPL et Javascript

Soit L une librairie javascript distribuée sous licence GPL . Soit E une entreprise qui modifie la librairie L pour en faire une version L2. L'entreprise E propose à des utilisateurs son application serveur qui utilise la version L2 de la librairie. Est ce que les utilisateurs peuvent exiger de récupérer le code source de la version L2 du logiciel ?

Oui
 Non

Give an explanation to your choice

En tant qu'utilisateurs de l'application serveur, on se retrouve dans le même cas que précédemment, on n'a pas de copie du logiciel. On ne peut donc toujours pas exiger récupérer le code.

What is the degree of confidence in your response (1: not confident to 5:very confident) ? 3

Submit

Second formulaire de soumission : une chance de modifier la première réponse

Reply Favorite I learn

FIGURE 2.4 – Tsaap-Notes : confrontation de point de vue et opportunité d'améliorer sa réponse

Seules les réponses correctes sont évaluées à l'aide d'une échelle de Likert à 5 niveaux. Chaque étudiant peut évaluer jusqu'à 3 réponses.

Focus sur l'algorithme d'appariement des réponses Durant la phase 2, chaque étudiant est confronté à une réponse proposée par un de ses pairs. Il a donc été nécessaire

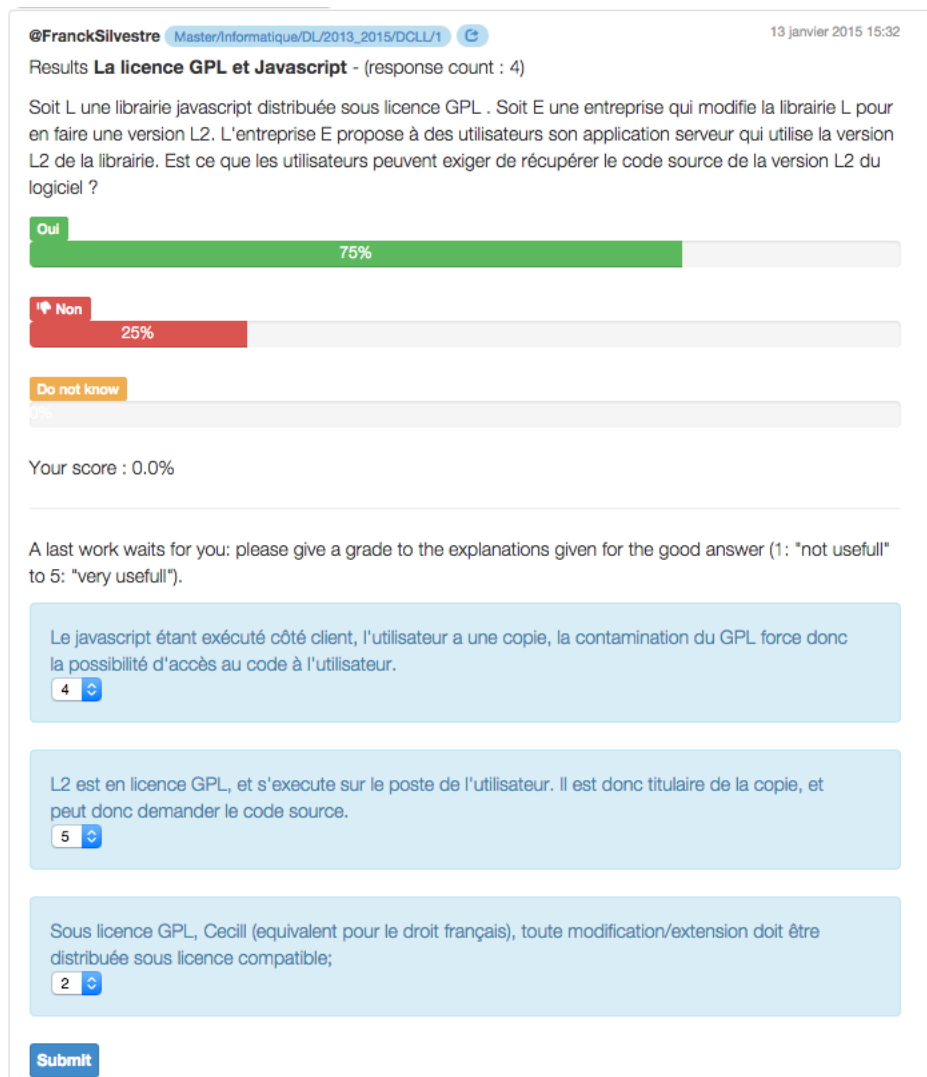


FIGURE 2.5 – Tsaap-Notes : histogramme des votes et évaluation par les pairs

d'implanter un algorithme d'appariement entre un étudiant et la réponse alternative la plus susceptible de générer un conflit cognitif favorable à l'apprentissage. Pour sélectionner une réponse candidate à une confrontation de point de vue, l'algorithme implanté dans Tsaap-Notes prend en compte la nature correcte ou incorrecte de la réponse, la taille de la réponse textuelle ainsi que le degré de confiance de l'étudiant ayant fourni la réponse. Pour chaque étudiant, la réponse de confrontation est sélectionnée suivant une approche *round robin* dans la liste des réponses différentes de la sienne candidates à une confrontation de point de vue. La figure 2.6 illustre l'algorithme : sur le graphique, les boîtes vertes et rouges représentent respectivement des réponses correctes et incorrectes. À l'issue de la première étape de l'algorithme, les listes des réponses candidates à la confrontation de point de vue sont constituées. C'est lors de la deuxième étape que les appariements en mode *round robin* sont effectués entre réponses correctes (respectivement incorrectes) et réponses candidates incorrectes (respectivement correctes).

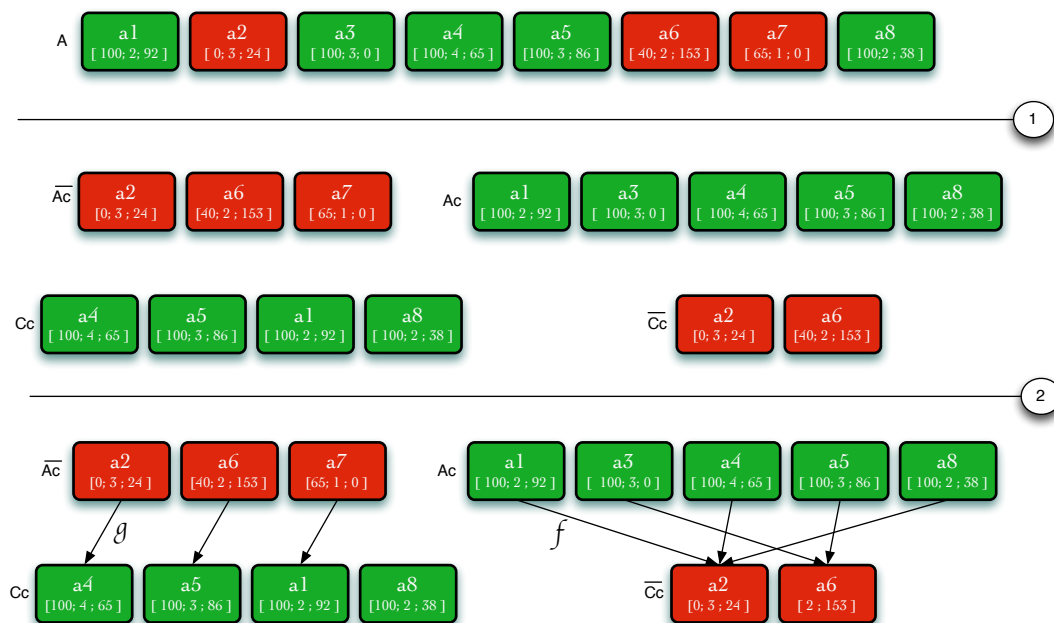


FIGURE 2.6 – Tsaaap-Notes : algorithme d'appariement des réponses pour la confrontation de point de vue

2.3.6 Expérimentations et résultats

La plateforme Tsaaap-Notes a été testée à l'Université Toulouse 3 Paul Sabatier dans le contexte de cours de première année de Master Informatique spécialité Développement Logiciel. Trois expérimentations ont été menées à différentes étapes des travaux de recherche durant trois années universitaires consécutives : 2012-2013, 2013-2014 et 2014-2015. Les groupes comprenaient entre 30 et 50 étudiants. Les résultats des deux premières expérimentations ont guidé la conception du processus en N Phase, préfigurant ainsi une approche *Design-Based Research* (ANDERSON & SHATTUCK, 2012) non encore mobilisée explicitement au cours de ces travaux.

Nous présentons ici le détail de la dernière expérimentation menée pour éprouver le processus en N Phases sur les hypothèses suivantes :

- (H1) Le processus engage fortement les étudiants sur l'activité de rédaction d'explications textuelles.
- (H2) La confrontation de point de vue entre pairs a un impact positif sur la quantité de bonnes réponses fournies à la deuxième soumission.
- (H3) L'évaluation par les pairs suscite un fort engagement des étudiants.
- (H4) Les étudiants perçoivent Tsaaap-Notes comme ayant une bonne utilisabilité.

L'expérimentation a été menée auprès de deux groupes d'étudiants dans le cadre de deux cours de Master 1 mention Informatique. Elle s'est déroulée sur 3 sessions de 2 heures pendant lesquelles 5 questions mobilisant le processus en N phases ont été posées aux étudiants.

Les résultats obtenus les plus saillants sont les suivants :

- 88% des étudiants ont fourni une réponse à la question posée, ce qui est en ligne avec les taux de participation mentionnés dans les autres études relatives aux systèmes de votes interactifs.
- 68,3% des étudiants ont fourni une réponse textuelle en complément de la sélection d'un item. Ce taux de participation, même s'il est en dessous du taux de participation à la fourniture d'une réponse, représente un bond quantitatif spectaculaire par rapport aux taux de participation constatés sur les tâches d'écriture avec les dispositifs testés ne proposant ni la confrontation de point de vue, ni l'évaluation par les pairs (au mieux, le taux de participation constaté à l'issue de la deuxième expérimentation était de 25%).
- En moyenne, 30% des étudiants ayant fourni une réponse erronée lors de la première phase ont répondu correctement lors de la seconde phase. Ceux-ci ont donc pleinement bénéficié de la confrontation de point de vue les ayant amenés à modifier leur croyance sur ce qui était vrai ou faux. Ce résultat est comparable aux résultats obtenus par SMITH et al. (2011) sur les études concernant l'« Instruction par les pairs ».
- 75,6% des étudiants ont participé à la tâche d'évaluation par les pairs. Cette tâche est proposée en phase 3 du processus et n'est pas obligatoire. Ce taux de participation révèle un engagement important sur cette activité mobilisant pleinement les pairs comme ressources d'instruction pour les autres en mobilisant les capacités métacognitives des étudiants.
- 43 étudiants ayant participé à l'expérimentation ont répondu au questionnaire *System Usability Scale*⁸ (BROOKE, 1996) permettant d'évaluer l'utilisabilité perçue d'un système par ses utilisateurs. Le score obtenu pour Tsaap-Notes est de 77,8/100. Il peut donc être considéré comme un système ayant une bonne utilisabilité perçue par les étudiants (BANGOR et al., 2009).

En synthèse, les résultats obtenus confirment les hypothèses formulées dans le cadre de cette expérimentation. Nous émettons cependant quelques réserves sur ces résultats :

- Le concepteur de la plateforme, l'enseignant et le chercheur ne faisaient qu'un : une telle concentration des responsabilités introduit très probablement des biais altérant les résultats.
- Le protocole d'expérimentation et les études statistiques menées sont sommaires.

Les résultats obtenus sont donc très encourageants, mais nécessitent d'être confirmés par des expérimentations et tests statistiques complémentaires.

8. <http://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html>

2.3.7 Discussion et positionnement des travaux

Le processus en N Phases introduit de nouvelles activités pour les étudiants et de nouveaux traitements à prendre en charge par le système technologique support afin de répondre à nos questions de recherche initiales.

QR 1.1 Lors de la première phase, les étudiants peuvent saisir une argumentation écrite lorsqu'ils répondent à une question fermée ou ouverte. Dans le système d'évaluation d'audience Tsaap-Notes, cette fonctionnalité exigée par le processus est implantée sous la forme d'un champ de saisie de texte proposé dans l'interface de réponse à une question. Le système collecte ainsi, en plus de la répartition des votes, les explications de chaque étudiant qui seront traitées dans les phases suivantes du processus.

QR 1.2 Lors de la deuxième phase, chaque étudiant consulte une réponse et une argumentation contraire à la sienne. Le système support à cette fonctionnalité fournit les réponses alternatives de manière anonyme pour limiter les biais d'influence relationnelle et exploite différentes variables caractérisant chaque réponse (caractère correct ou incorrect, degré de confiance de l'auteur, taille de la réponse textuelle) pour présenter à chaque étudiant une réponse alternative. L'objectif est de provoquer un conflit sociocognitif dans des conditions le rendant propice à l'apprentissage (BUCHS et al., 2008) Cette deuxième phase offre ensuite l'opportunité à chaque étudiant de modifier sa réponse. Enfin, la dernière phase du processus consiste à faire évaluer, par les étudiants, les réponses correctes produites par les autres. Cette activité vise d'une part à évaluer dans un temps raisonnable les contributions textuelles des étudiants et, d'autre part à qualifier les meilleures réponses d'étudiants en vue de recycler ces réponses dans des tests de révision (cf. chapitre 3). Les phases 2 et 3, en mobilisant fortement les contributions des pairs (fourniture de réponses contradictoires en phase 2, fournitures d'évaluations en phase 3), s'inscrivent pleinement dans la 5^e stratégie caractéristique de l'évaluation formative (BLACK & WILIAM, 2009) consistant à utiliser les pairs comme ressources d'instruction pour les apprenants. Les résultats de l'expérimentation exposée précédemment indique des progressions entre le premier et le deuxième vote relativement similaires à celles constatées dans les études sur le *Peer Instruction*. Ces résultats, avec toutes les réserves sur la qualité de l'expérimentation, encouragent à penser que le processus en N Phases et son implantation dans Tsaap-Notes, n'introduisent pas de régression dans les bénéfices produits par le *Peer Instruction*.

QR 1 Comme indiqué en sous-section 2.3.3, nous avons élaboré une grille permettant d'évaluer et comparer les processus d'évaluation d'audience. Cette grille a guidé la conception du processus en N Phases ainsi que son implantation dans Tsaap-Notes. Le processus proposé dépasse donc par construction les limites identifiées dans les processus étudiés. Le

tableau 2.2 présente l'évaluation du processus en N Phases implanté dans Tsaap-Notes en comparaison avec les processus étudiés durant la thèse. Dans ce tableau, les sigles SVI et SEPT signifient respectivement *Systèmes de Votes Interactifs* et *Systèmes d'échanges et de partages de textes* et caractérisent les familles de processus étudiés durant la thèse. La tableau montre que le processus implanté dans Tsaap-Notes est plus qu'une juxtaposition des fonctionnalités des processus étudiés : il introduit des activités caractéristiques de l'évaluation formative non couvertes par les processus SVI et SEPT en introduisant par exemple l'évaluation par les pairs.

TABLE 2.2 – Évaluation de la plateforme Tsaap-Notes

	Activités de référence					Stratégies						Interactions		Eng.
	A-1	A-2	A-3	A-4	A-5	S-1	S-2.1	S-2.2	S-3	S-4	S-5	I-ST	I-FT	E-TE
SVI	■	■				■	■	■	■	■	■	■	■	■
SEPT							■	■	■	■	■	■	■	■
Tsaap-Notes	■	■	■	■	■	■	■	■	■	■	■	■	■	■

Positionnement par rapport à l'échelle ICAP Les travaux de ma thèse n'ont pas exploré les liens entre les processus d'évaluation d'audience et le cadre ICAP permettant de lier les niveaux d'engagement cognitifs (Interactif > Constructif > Actif > Passif) à des activités pédagogiques concrètes améliorant les apprentissages (CHI et al., 2018 ; CHI & WYLIE, 2014). Je n'avais tout simplement pas connaissance de ces travaux en 2015. Dès la lecture des travaux de CHI et WYLIE, il est apparu intéressant de positionner les processus d'évaluation d'audience au regard des niveaux d'engagements ICAP. Nous avons proposé en 2019 une couverture des niveaux d'engagement dans le *Peer Instruction* et dans le processus implanté dans la plateforme Elaastic (SILVESTRE, 2019). Il ressort de ce travail analytique que le *Peer Instruction*, à travers l'introduction des micro-débats après la première phase de réponse à une question, engage les étudiants sur le niveau Interactif, le plus haut niveau d'engagement dans l'échelle ICAP. Dans le processus Elaastic, l'élaboration de l'argumentation écrite justifiant le choix de la réponse engage les étudiants sur le niveau Constructif. Il subsiste un doute sur le niveau d'engagement induit par la séquence de confrontation de points de vue anonymisée s'appuyant uniquement sur l'évaluation de réponses alternatives. En effet, bien que les productions des pairs soient mobilisées dans cette phase du processus, il n'y a pas d'interactions directes entre les étudiants, pas d'échange ni de discussion. Il ne peut donc pas être affirmé que le niveau d'engagement produit soit équivalent au niveau Interactif. Nous souhaitons dans les travaux futurs éclaircir ce point à l'aide d'expérimentations permettant de comparer les résultats obtenus sur les différents processus.

Positionnement par rapport aux activités pédagogiques de type *Think Pair Share* Une activité *Think Pair Share* (TPS) est une activité pédagogique coopérative

pouvant être proposée dans des classes de différentes tailles dans n'importe qu'elle discipline enseignée. L'enseignant pose une question ou un problème sur laquelle ou lequel les étudiants commencent par réfléchir de manière individuelle (phase *Think*). Ils sont ensuite invités à discuter de leur réponse avec un ou une voisine en vue de produire une réponse commune (phase *Pair*). L'activité se termine par une discussion d'ensemble où les réponses obtenues en petits groupes sont exposées à la classe (phase *Share*) (LIGHTNER & TOMASWICK, 2017). L'approche TPS a été décrite pour la première fois par LYMAN (1981) et a fait l'objet de diverses études démontrant les bénéfices de l'approche sur l'engagement de tous les apprenants participant à une telle activité ainsi que sur les résultats d'apprentissages (APRIYANTI & AYU, 2020 ; FITZGERALD, 2013 ; KADDOURA, 2013). KOTHIYAL et al. (2014) analysent finement les points communs et les différences entre l'approche TPS et le *Peer Instruction* (PI) : les premières phases dans le TPS sont clairement orientées sur la construction puis la coconstruction d'une réponse à une question ou à un problème alors que les deux premières phases dans le *Peer Instruction* sont centrées sur la sélection d'une réponse dans une question à choix et sur l'argumentation en micro-débat concernant cette sélection. Dans son manuscrit d'HDR, BROISIN (2020) positionne directement les travaux réalisés pendant ma thèse avec l'approche TPS (BROISIN, 2020). Le processus en N Phases est ainsi présenté comme une déclinaison de processus TPS dans le contexte d'un enseignement à très large audience. La première phase du processus en N Phases se rapproche effectivement de la première phase du TPS en proposant une activité générative sur la réponse à la question posée. Cependant, la deuxième phase du processus en N Phases, bien que mobilisant les productions des pairs, n'intègre pas la dimension coconstruction d'une réponse caractéristique du TPS. De la même manière qu'il nous a paru intéressant de comparer les résultats obtenus avec le PI et le processus en N Phases (cf. paragraphe précédent), il nous semble pertinent de comparer les résultats obtenus avec une approche rigoureusement TPS et le processus en N Phases. L'enjeu de ces futures études, *in fine*, est de savoir s'il est aussi efficace d'engager les étudiants dans un dialogue en face à face que d'organiser une confrontation de points de vue anonyme et médiatisée par la technologie pouvant opérer dans différents contextes d'enseignements (hybride ou à distance).

Publications Les travaux présentés dans cette section ont fait l'objet de plusieurs publications. Les modèles BPMN de processus et la grille d'évaluation ont été principalement présentés dans le manuscrit de ma thèse (SILVESTRE, 2015). Le processus en N Phases et les résultats associés concernant d'une part l'engagement élevé des étudiants durant les séquences et d'autre part les taux de changement de réponses ont été publiés à la conférence EC-TEL 2015 (SILVESTRE et al., 2015c) ainsi que dans la revue STICEF (SILVESTRE et al., 2017). Les étapes de construction du processus en N Phases partant initialement d'une approche visant à engager les étudiants dans une activité de prise de notes collaborative

ont fait l'objet de plusieurs publications entre 2014 et 2015 (SILVESTRE et al., 2014a, 2014b, 2015a, 2015b).

2.3.8 Rétrospective

Nous proposons une rétrospective de cette première tranche de travaux s'inspirant des rétrospectives conduites dans les projets informatiques menés suivant une méthode agile. Cette démarche sera déclinée pour l'ensemble des travaux décrits dans ce document. Nous forçons ainsi une posture réflexive sur le processus macroscopique de recherche, à l'échelle d'un travail conduisant à l'obtention de l'HDR, en plus de celle apportée par la discussion sur les contributions.

Les forces

Des recherches fondements d'autres travaux Les travaux visant à répondre à la première question de recherche ont été principalement réalisés dans le cadre de ma thèse. Ces travaux ont posé les bases de ce qui me permet de rédiger ce document aujourd'hui : ils ont été le socle des recherches qui ont suivi. Ils sont encore le socle du projet de recherche décrit dans la partie II.

Des résultats dignes d'intérêt pour les praticiens Le système Tsaap-Notes a évolué vers un nouveau système, baptisé *Elaastic*, implantant un processus légèrement différent du processus en N Phases tout en conservant ses fondements : argumentation écrite, évaluation par les pairs et possibilité de changer sa réponse. Le système Elaastic fera l'objet d'une description détaillée dans la section 2.4. Ces deux systèmes ont été conçus dès l'origine pour un usage en contexte écologique, et ont été testés avec de « vrais » étudiants. « *Eat your own dog food* » disent les anglo-saxons. Si la démarche d'effectuer ses premières expérimentations en tant que chercheur et concepteur des solutions proposées soulèvent des questions légitimes sur les biais induits, elle permet de démontrer que les systèmes proposés sont utilisables en situations réelles.

Ces deux systèmes ont suscité et suscitent encore l'intérêt des praticiens, c'est-à-dire des enseignants du supérieur comme du secondaire qui voient dans ces plateformes des outils utiles et cohérents avec leurs pratiques pédagogiques. Cet intérêt a permis de monter des projets de recherches collaboratives avec les académies de Nancy-Metz et de Toulouse. La thèse de Rialy Andriamiseza a été financée par la Direction du Numérique Éducatif dans le cadre d'une collaboration avec l'Académie de Nancy-Metz.

L'environnement technique comme capital Enfin, Tsaap-Notes et Elaastic ont été conçus pour être robustes, maintenables et fiables. Notre objectif a toujours été de proposer des solutions pouvant encaisser la charge d'un déploiement en contexte écologique

sur plusieurs établissements tout en étant capable d'évoluer à moindre coût. Elaastic, plateforme sur laquelle s'implantent aujourd'hui les évolutions issues de nos travaux, offre un environnement technique sur lequel nous sommes en mesure d'intégrer les nouvelles évolutions en capitalisant sur les développements initiés en 2015 (après refonte de Tsaap-Notes).

Les faiblesses

Un ancrage théorique trop tardif Les travaux menés autour de la conception de systèmes d'évaluation d'audience ont été menés sur l'ensemble de ma thèse de manière très exploratoire, trop exploratoire. Les travaux portaient initialement sur la manière d'engager les étudiants dans des activités d'écritures en amphithéâtre. Est apparue l'idée de recycler les productions pour faciliter la génération de tests de révisions. J'avais alors à tâtons en résolvant des micro-problèmes peinant à s'inscrire dans une vision globale. Chaque papier faisait l'objet d'une micro revue de la littérature relative à chaque micro-problème. Approche suffisante pour l'écriture d'un papier de conférence mais insuffisante pour un manuscrit de thèse. De fait, l'écriture du manuscrit a été une ré-écriture complète de l'histoire, une sorte de *reverse researching*, ayant consisté, à partir des résultats obtenus et publiés, à reconstruire une histoire scientifique cohérente. Cette histoire scientifique a nécessité en fin de thèse une revue de littérature complémentaire importante. Les travaux de la thèse ne se sont ancrés dans le cadre de l'évaluation formative qu'à l'issue de cette revue de littérature complémentaire. Comme cela était déjà indiqué dans le « Précis de recherche en Ingénierie des EIAH » (TCHOUNIKINE, 2009), l'ancrage au plus tôt dans un cadre théorique robuste issu des sciences de l'éducation et/ou de la psychologie est indispensable pour faciliter le travail du doctorant durant sa thèse et plus spécifiquement au moment de la rédaction du manuscrit.

Le nommage des concepts et des systèmes Processus en N Phases, Tsaap-notes, etc. Autant de noms qui n'ont pas survécu au travail réalisé pendant la thèse. Les noms que nous donnons aux nouveaux concepts, aux nouveaux processus ou logiciels constituent aussi un capital pour les travaux futurs, à condition d'être bien choisis. Ce ne fut pas le cas des productions de ma thèse.

Le passage de Tsaap-Notes à Elaastic mal documenté Le processus en N Phases présentés durant la thèse a évolué dès 2015 vers un processus différent présenté dans l'article « La (dé-)synchronisation des transitions dans un processus d'évaluation formative exécuté à distance : impact sur l'engagement des étudiants » de PARMENTIER et SILVESTRE (2019). La justification du changement de processus n'a pas fait l'objet de publication officielle : le papier coécrit avec Jean-François Parmentier présentant les arguments de l'évolution et montrant l'absence d'impact sur les résultats de séquences

comparables jouées avec l'un ou l'autre des processus a fait l'objet d'un rejet à la conférence LAK. Les résultats n'étaient pas remis en cause, mais la contribution du papier, très mal mise en valeur dans notre proposition, n'a pas été perçue comme suffisamment significative pour faire l'objet d'une publication. Nous n'avons pas pris le temps d'améliorer l'article soumis. À défaut, notre soumission est disponible sur HAL (SILVESTRE & PARMENTIER, 2021) dans la catégorie *Document de travail*. Plus globalement, cette lacune dans ces recherches initiales nous incite à documenter au fil de l'eau les travaux dans un esprit de sciences ouvertes. La mise en place de projets sur la plateforme *Open Science Framework*⁹ vise à améliorer cet axe de notre processus de recherche.

Des expérimentations très perfectibles La plupart des expérimentations dans le cadre de ma thèse ont été menées dans le cadre de mes cours auprès de mes étudiants. Une telle approche induit inévitablement des biais que nous souhaitons éviter dans les travaux futurs. Mes compétences en statistiques et plus généralement en analyses de données sont limitées. Les analyses réalisées sur les données collectées durant les travaux de ma thèse sont donc superficielles. Ce déficit de compétence devra être comblé d'une manière ou d'une autre dans les travaux à venir.

Manque de ressources humaines sur le projet de thèse J'ai effectué ma thèse dans des conditions particulières : en poste d'enseignant-chercheur à mi-temps, je ne disposais en moyenne que d'un tiers temps pour travailler sur mon sujet de recherche. À ce temps de travail s'est ajouté le temps passé par Julien Broisin dans le cadre de son activité de direction de la thèse (Philippe Vidal n'a malheureusement pas pu pleinement assurer cette fonction en raison de problèmes de santé). En cumulant les temps passés par les deux personnes impliquées sur la thèse, nous évaluons l'effort en ressources humaines sur les 3 ans et 3 mois à moins de 2 années-homme. Cela nous paraît bien trop peu de temps consacré pour mener des travaux de recherche quantitativement significatifs et de bonne qualité.

2.4 Compréhension et amélioration des processus à 2 votes

2.4.1 Méthode de conduite de la recherche

L'approche proposée dans la thèse de Rialy est de type *Design Based Research* (DBR). Deux itérations ont constitué le cycle DBR mis en œuvre. La première itération, initiée par une analyse de données exploratoire, a conduit à la proposition d'un nouveau modèle

9. <https://osf.io/>

d'orchestration s'appliquant aux processus à deux votes. La deuxième itération a porté sur la mise à l'épreuve de ce nouveau modèle en contexte écologique.

Cette recherche de type DBR mobilise l'analyse des données d'apprentissage dans le but d'améliorer la compréhension et la mise en œuvre de séquences d'évaluation formative. À ce titre, elle s'inscrit dans le champ des *Learning Analytics* (LA) (SOLARIS, 2012). Le cycle des *Learning Analytics* (CLOW, 2012) présenté dans la figure 2.7 vise la fourniture d'un cadre facilitant la conduite de travaux relevant du champ des LA. Sa nature cyclique le rend compatible avec une approche DBR et permet de synthétiser la nature des itérations opérées durant la thèse de Rialy.

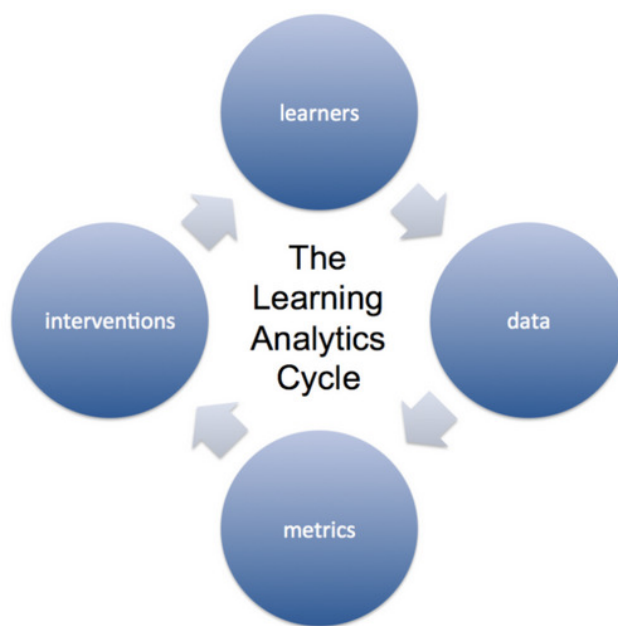


FIGURE 2.7 – Le cycle des *Learning Analytics* (CLOW, 2012)

2.4.2 L'analyse de données exploratoire

Le processus à 2 votes dans Elaastic

Elaastic est une application dérivée de Tsaap-Notes pour répondre à de nouveaux besoins : réduire la durée des séquences et permettre aux enseignants d'obtenir plus d'informations sur les justifications des choix erronés des étudiants. Ainsi, la deuxième phase du processus Tsaap-Notes a été remaniée. Chaque étudiant utilise une échelle de Likert à 5 niveaux pour évaluer son degré d'accord avec un ensemble de contributions fournies par d'autres étudiants. Cet ensemble peut contenir jusqu'à 5 arguments différents et inclut des explications pour les réponses correctes et incorrectes. La figure 2.8 présente l'interface de la phase 2 implantée dans Elaastic.

Contrairement à Tsaap-Notes, les justifications présentées ne sont plus sélectionnées

FIGURE 2.8 – Elaastic - La phase 2 composée d'une interface pour l'évaluation par les pairs de réponses correctes et incorrectes.

en fonction du degré de confiance de leur auteur, assurant ainsi que chaque justification reçoive un nombre significatif d'évaluations. Le nombre de contributions évaluées par chaque étudiant est défini par l'enseignant au début de la séquence.

Le processus à 2 votes implanté dans Elaastic présente deux avantages principaux :

- il est condensé sur 3 phases, contre 4 dans le processus Tsaap-Notes ;
- il fournit aux enseignants des évaluations faites par les étudiants sur toutes les justifications, qu'elles soient correctes ou incorrectes.

Le Tableau 2.3 résume les différentes phases des deux processus, mettant en évidence les différences entre Tsaap-Notes et Elaastic, notamment en ce qui concerne l'évaluation

des contributions par les étudiants.

TABLE 2.3 – Comparaison entre les phases dans Tsaap-Notes et Elaastic

	Tsaap-Notes	Elaastic
Phase 1	Chaque étudiant répond à une question en fournissant une justification écrite et un degré de confiance	
Phase 2	Chaque étudiant consulte exactement une justification fournie par un pair et correspondant à une réponse différente.	Chaque étudiant évalue jusqu'à cinq justifications fournies par des pairs et correspondant à des réponses correctes ou incorrectes.
	Chaque étudiant soumet une nouvelle réponse avec la possibilité de changer le choix initial et l'argumentation.	Chaque étudiant soumet une nouvelle réponse avec la possibilité de changer le choix initial.
Phase 3	Chaque étudiant évalue 3 argumentations correspondant uniquement à la réponse correcte.	-
Dernière Phase	Restitution des résultats et discussion	

Le document de travail rédigé par SILVESTRE et PARMENTIER (2021) montre que, sur des séquences comparables, la différence entre les tailles d'effet des séquences orchestrées avec Tsaap-Notes et celles orchestrées avec Elaastic n'est pas significative.

Description du jeu de données

Le jeu de données utilisé pour l'analyse exploratoire est constitué de données collectées dans le cadre de l'utilisation d'Elaastic dans l'enseignement supérieur de 2015 à 2019. Pendant cette période, 623 séquences ont été collectées auprès de 53 enseignants, avec la participation de 1 769 apprenants qui ont fourni 8 757 réponses et effectué 9 256 évaluations par les pairs. Le jeu de données est structuré autour de *séquences* qui comprennent divers éléments tels que le contexte d'exécution, le nombre de réponses devant être évaluées par les pairs et le nombre de participants. Chaque séquence contient aussi une question, un ensemble de réponses et un ensemble d'évaluations. Pour chaque réponse, des informations comme le contenu de la justification, le score et les choix sélectionnés sont recueillis.

Le jeu de données initial a été filtré pour répondre aux besoins de l'analyse exploratoire. Seules les séquences portant sur des questions à choix et exécutées en face à face ont été retenues. Celles où le nombre de participants était inférieur à 10 ont été éliminées. Après l'application de ces filtres, l'échantillon final compte 104 séquences réalisées par 21 enseignants, où 616 apprenants ont fourni 1 981 réponses et réalisé 4 072 évaluations par les pairs. Ces séquences portent principalement sur des sujets STEM (Science, Technologie, Ingénierie et Mathématiques) dans l'enseignement supérieur. Dans la suite du document, nous désignons cet échantillon par l'expression *Échantillon 2019*.

Taille d'effet d'une séquence

Afin de mesurer l'impact bénéfique d'une séquence d'enseignement, nous avons considéré deux variables : p_1 et respectivement p_2 , qui représentent la proportion d'apprenants ayant répondu correctement lors du premier vote et, respectivement, lors du second vote.

Une séquence est alors qualifiée de *bénéfique* si $p2 > p1$, de *non bénéfique* si $p2 = p1$, et de *nuisible* si $p2 < p1$.

Cependant, cette classification ne permet pas de différencier une séquence où $p1 = 10\%$ et $p2 = 11\%$ d'une séquence où $p1 = 10\%$ et $p2 = 90\%$: la taille de l'effet n'est pas prise en compte dans cette classification rudimentaire. PARMENTIER (2018) propose une estimation de la taille de l'effet de Cohen (égalité 2.1) permettant de fournir une meilleure évaluation de l'impact d'une séquence.

$$d = 0.6 \ln \left(\frac{p2}{1 - p2} \frac{1 - p1}{p1} \right) \quad (2.1)$$

La figure 2.9 illustre la répartition des séquences de l'échantillon 2019 par rapport à leur taille d'effet. Sur les 104 séquences, 65 séquences ont été bénéfiques (représentées par les points bleus), 27 non bénéfiques (représentées par les points verts) et 12 nuisibles (représentées par les points rouges). Face à ce constat, le travail de l'analyse exploratoire a consisté essentiellement à déterminer les variables influençant la taille d'effet d'une séquence. Ce travail a conduit à la formulation de recommandations à destination des enseignants et des concepteurs de systèmes. Elles visent à augmenter le ratio des séquences bénéfiques par rapport aux séquences non bénéfiques ou nuisibles.

2.4.3 Résultats et recommandations induites

Les premiers résultats mis en lumière par l'étude exploratoire concernent les indicateurs permettant de qualifier la qualité d'une séquence d'évaluation formative orchestrée suivant un processus à 2 votes :

- Plus le taux de bonnes réponses au premier vote est proche de 50%, plus les chances d'obtenir une séquence bénéfique augmente.
- Les séquences ont plus de chance d'être bénéfiques quand les apprenants évaluent de manière cohérente les justifications de leurs pairs, c'est-à-dire lorsqu'ils donnent des notes plus élevées aux justifications pour les bonnes réponses et des notes plus faibles aux justifications pour les mauvaises réponses.
- Les apprenants ont plus de chance d'évaluer de manière cohérente les justifications de leurs pairs lorsque les degrés de confiance sont cohérents à l'issue du premier vote, c'est-à-dire quand les apprenants ayant fourni une bonne réponse ont indiqué un degré de confiance élevé alors que ceux qui ont fourni une mauvaise réponse ont indiqué un degré de confiance faible.
- Cependant, les chances d'obtenir une séquence bénéfique n'augmentent pas significativement lorsque les degrés de confiance sont cohérents.
- L'autoévaluation explicite produit des résultats biaisés lorsqu'elle est proposée dans le contexte d'évaluations par les pairs.

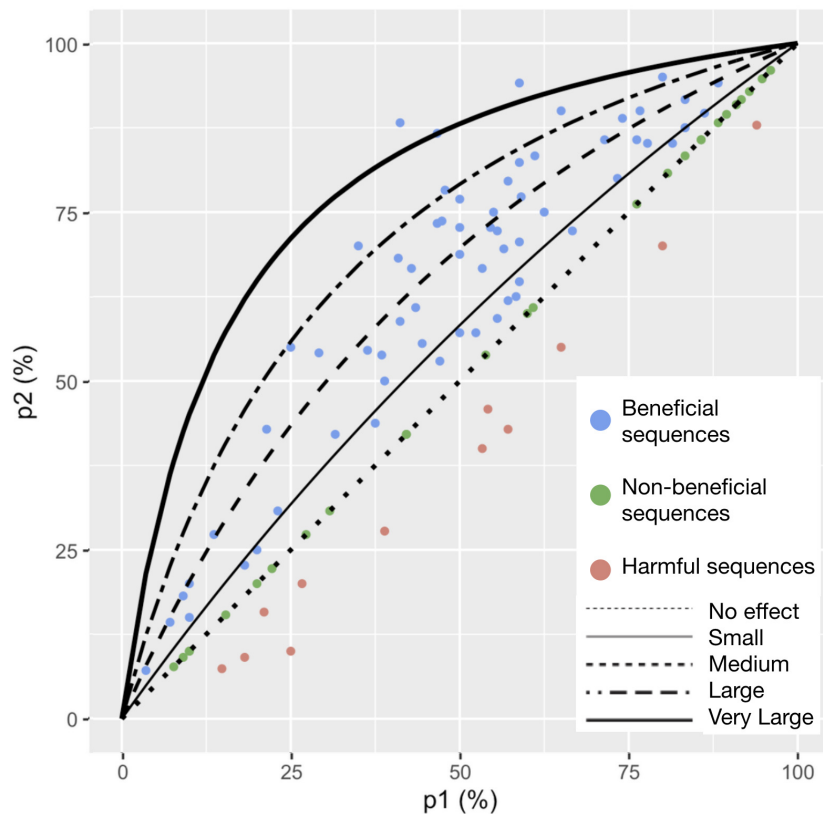


FIGURE 2.9 – Échantillon 2019 - Séquences et lignes de la taille de l'effet de Cohen d : aucun effet ($d = 0$), petit ($d = 0,2$), moyen ($d = 0,5$), grand ($d = 0,8$) et très grand ($d = 1,2$).

- Le nombre d'évaluations que chaque apprenant effectue au sein des activités d'évaluation par les pairs ne fait pas de différence significative en termes d'effets sur les séquences.

Un ensemble de recommandations pour les concepteurs de systèmes d'évaluation formative implantant un processus à 2 votes ont été formulées sur la base de ces résultats.

RCS-1 : Les systèmes d'évaluation formative implantant un processus à 2 votes devraient fournir aux enseignants une indication sur la cohérence des degrés de confiance des apprenants à l'issue du premier vote. Ils devraient également introduire de la flexibilité concernant la manière de poursuivre la séquence en fonction de cette indication de cohérence.

RCS-2 : Les systèmes d'évaluation formative mettant en œuvre une activité d'évaluation par les pairs devraient fournir aux enseignants une indication sur la cohérence des évaluations et introduire de la flexibilité concernant la manière de poursuivre la séquence en fonction de cette indication de cohérence.

RCS-3 : Les systèmes d'évaluation formative implantant un processus à 2 votes devraient introduire de la flexibilité concernant la sélection des justifications au cœur de la discussion, en fonction de la cohérence des degrés de confiance.

RCS-4 : Les activités d'évaluation par les pairs dans les systèmes d'évaluation formative ne devraient pas inclure d'autoévaluation.

RC-5 : Les systèmes d'évaluation formative devraient introduire de la flexibilité concernant le nombre de réponses à évaluer par les apprenants.

Des recommandations d'orchestration de séquences destinées aux enseignants ont été formulées sur la base des mêmes résultats.

RO-1 : Après le second vote, les explications fournies par les enseignants devraient être plus détaillées si la séquence n'a pas été bénéfique. ❶

RO-2 : À la fin du premier vote :

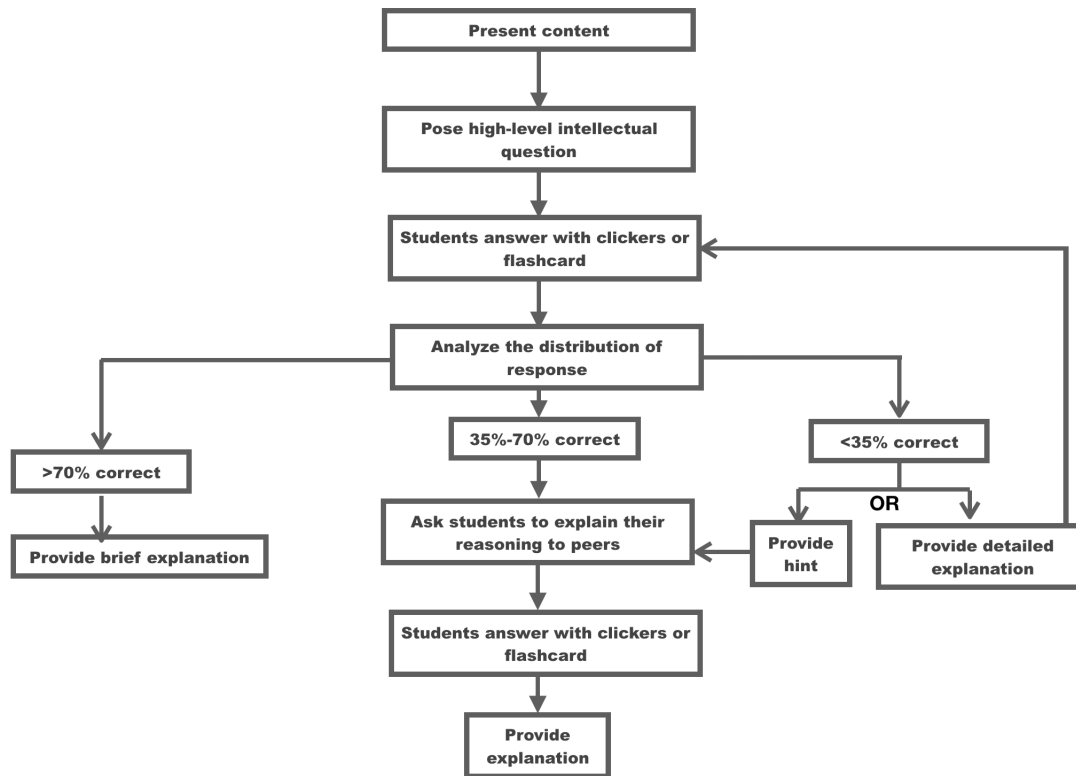
- Si plus de 80% des réponses sont correctes, les enseignants devraient sauter la phase de confrontation. ❷
- Lorsqu'il y a moins de 30% de bonnes réponses :
 - Si les degrés de confiance sont cohérents, les enseignants devraient donner des indices aux apprenants avant de commencer la phase de confrontation.
 - Sinon, les enseignants devraient fournir des explications détaillées et recommencer une nouvelle séquence pour évaluer le même concept.

❸

RO-3 : Si les mauvaises réponses sont mieux notées que les autres, les enseignants devraient se concentrer sur ces réponses lors de la discussion. Sinon, les enseignants devraient se concentrer sur les bonnes réponses lors de la discussion. ❹

RO-4 : Lorsque les étapes de confrontation et de second vote ont été sautées durant la séquence :

- Si les degrés de confiance sont incohérents, les enseignants devraient se concentrer sur les justifications des mauvaises réponses lors de la discussion.
- Sinon, les enseignants devraient se concentrer sur les justifications des

FIGURE 2.10 – Modèle d’orchestration de Vickrey élaboré pour le *Peer Instruction*

bonnes réponses lors de la discussion. 5

RO-5 : Les enseignants peuvent décider du nombre de contributions à évaluer par chaque apprenant dans la phase de confrontation de points de vue.

2.4.4 Un nouveau processus à 2 votes

Le modèle d’orchestration

Les recommandations d’orchestration ont conduit Rialy à formaliser un modèle d’orchestration de processus à 2 votes dérivé de celui proposé par VICKREY et al. (2015) pour le *Peer Instruction* (figure 2.10). Notons que ce dernier n’est pas entièrement déterministe : dans le cas d’un taux de réponses inférieur à 35%, le choix de poursuivre ou non la séquence est laissé à la discrétion de l’enseignant sans qu’aucun indicateur puisse éclairer la décision.

Le modèle proposé par Rialy (figure 2.11) complète le modèle de VICKREY et al. sur la base des résultats de l’analyse exploratoire et est entièrement déterministe. Sur la figure 2.11, les numéros indiqués sur les nœuds de décision renvoient directement aux numéros indiqués dans les recommandations d’orchestration.

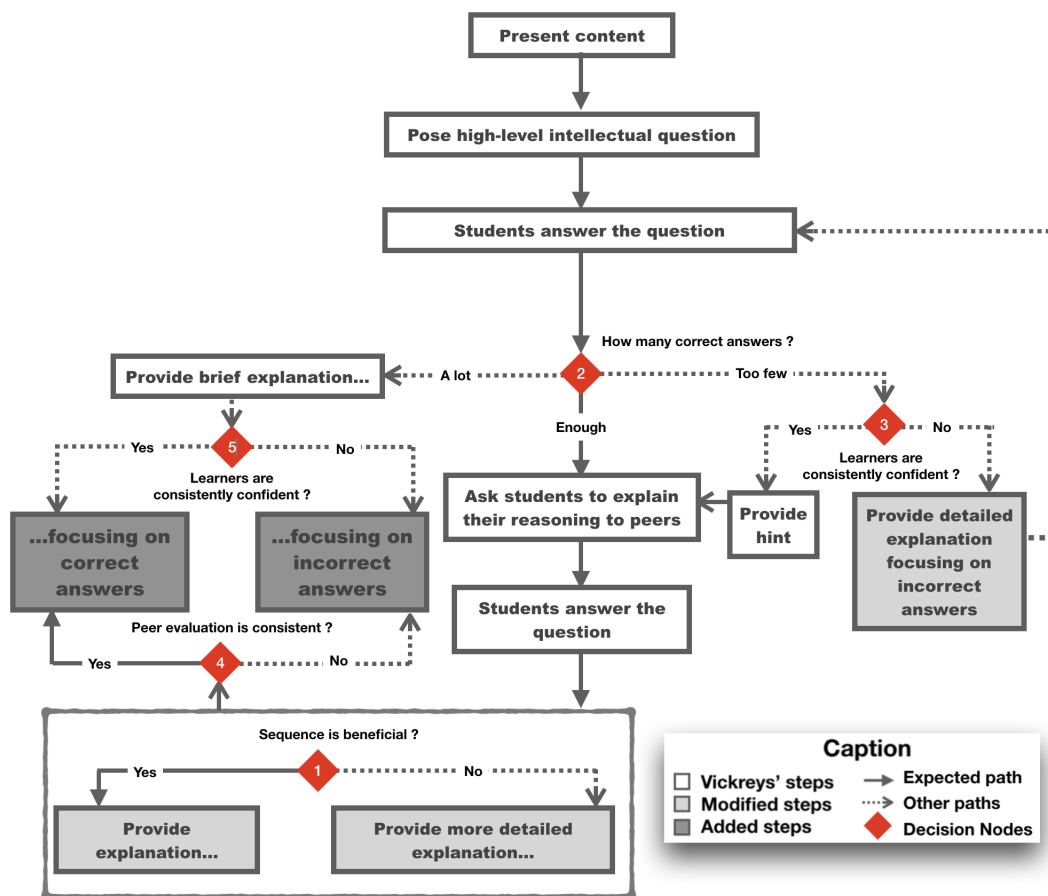


FIGURE 2.11 – Nouveau modèle d'orchestration pour processus à 2 votes

L'implantation dans Elastic

Le modèle d'orchestration présenté dans la section précédente a été implanté dans Elastic afin de pouvoir l'éprouver en contexte écologique. Le parti pris a été de toujours laisser le choix à l'enseignant de conduire la séquence selon sa volonté. L'implantation dans Elastic se contente donc de présenter les indicateurs à l'enseignant et de lui proposer sous forme de recommandation la décision indiquée par le modèle.

Un travail complémentaire de revue de littérature dans le champ de l'explicabilité des algorithmes d'intelligences artificielles (CHENG et al., 2019; MILLER, 2019; RIBERA & LAPEDRIZA GARCIA, 2019; SCHAFER et al., 1999; TINTAREV & MASTHOFF, 2015) a influencé la manière de présenter et de justifier les recommandations. Ainsi, les explications de chaque recommandation sont accessibles à la demande et suivant plusieurs niveaux de détails. La figure 2.12 illustre un exemple de présentation de recommandation. L'accès aux explications est conditionné par le clic de l'enseignant sur le lien *Read more*. La figure 2.13 présente la fenêtre apparaissant quand l'enseignant clique effectivement sur ce lien. Il peut alors accéder à un niveau de détail supplémentaire en cliquant sur le petit triangle correspondant à une explication. Un diagramme peut alors apparaître pour compléter l'explication comme illustré sur la figure 2.14.

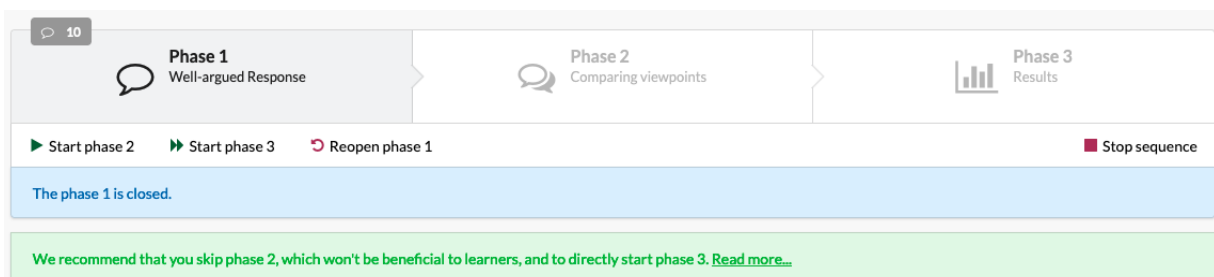


FIGURE 2.12 – Elastic - Explication de premier niveau

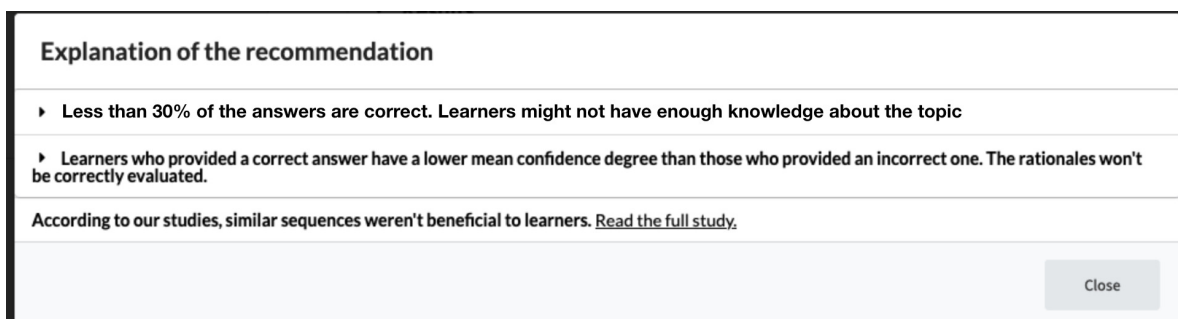


FIGURE 2.13 – Elastic - Explication de niveau 2

La restitution des résultats a été modifiée pour mettre en valeur les explications à discuter suivant les recommandations du modèle d'orchestration. La figure 2.15 présente l'interface de présentation des résultats incluant le nouvel ensemble d'explications.

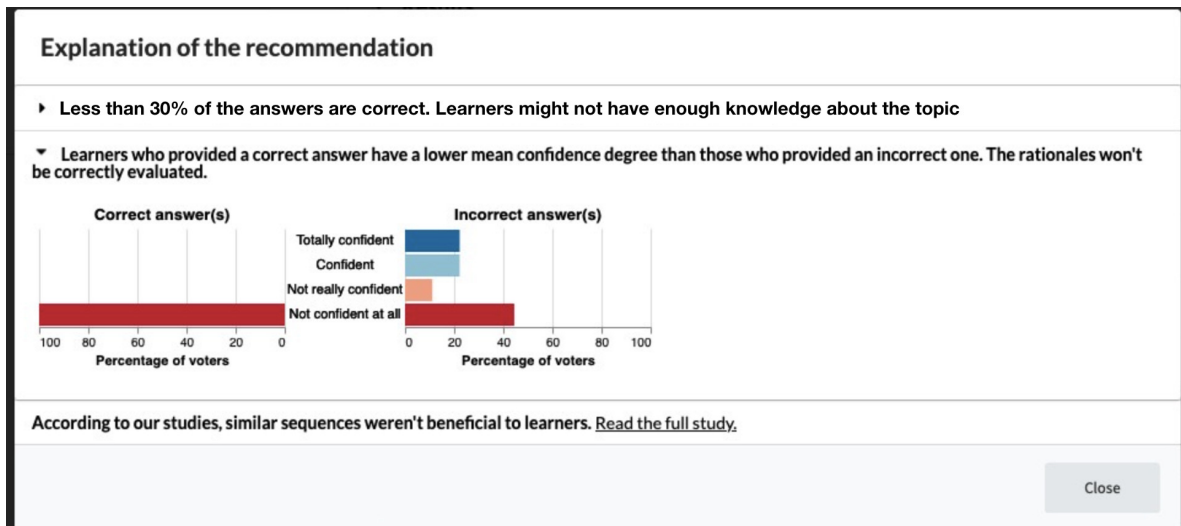


FIGURE 2.14 – Elaastic - Explication avec haut niveau de détail

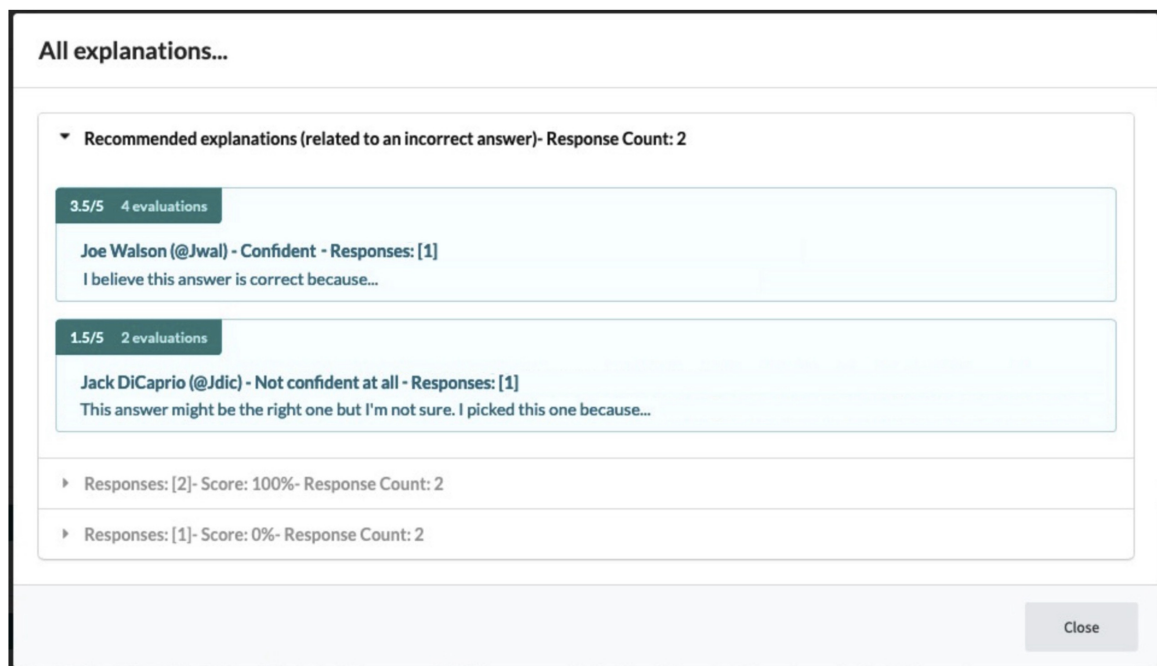


FIGURE 2.15 – Elaastic - Recommandations sur les explications à discuter en priorité

Afin de mesurer le suivi des recommandations par les enseignants, une série de sondes logicielles ont été implantées dans Elaastic. La version 5.0 d'Elaastic implantant cet ensemble de nouvelles fonctionnalités a été déployée en février 2022 et a permis de collecter des traces sur une période 5 mois.

2.4.5 Expérimentation et résultats

L'expérimentation menée visait à mettre à l'épreuve les hypothèses suivantes :

(H1) Les enseignants suivent les recommandations induites par le nouveau modèle d'or-

chestration.

(H2) Sur le nombre de séquences complètes, c'est-à-dire, durant lesquelles les apprenants ont participé aux 3 phases du processus à 2 votes implanté dans Elaastic, le taux de séquences bénéfiques est significativement supérieur au taux observé lors de l'étude exploratoire.

Pour l'hypothèse (H2), le pourcentage de séquences bénéfiques attendu a été évalué en étudiant les séquences de l'étude exploratoire se déroulant conformément au modèle d'orchestration proposé par Rialy. La valeur attendue fut évaluée à 76%.

Description du jeu de données

Les données collectées concernent 212 séquences orchestrées par 19 enseignants en collège durant lesquelles 584 élèves ont fourni 2629 réponses et produit 23637 évaluations par les pairs. Ce jeu de données a été filtré en appliquant les critères utilisés lors de constitution de l'échantillon de l'étude exploratoire (cf. section 2.4.2). L'échantillon obtenu au final concerne 118 séquences orchestrées par 5 enseignants durant lesquelles 436 élèves ont fourni 1988 réponses et produit 6036 évaluations par les pairs.

Résultats et analyse

(H1) Les enseignants ne suivent pas les recommandations L'analyse des données issues des sondes traçant le suivi des recommandations est sans appel : les enseignants ne suivent pas les recommandations. Une étude qualitative informelle a alors été menée auprès de 3 enseignants pour tenter de déterminer les causes de cette attitude vis-à-vis des recommandations. Les réponses fournies concernent exclusivement un problème majeur d'utilisabilité : les enseignants, pris dans le feu de l'action de l'orchestration de la séquence, ne voient tout simplement pas les recommandations.

(H2) Le taux de séquences bénéfiques est très en dessous du taux attendu La figure 2.16 résume la répartition des séquences étudiées en fonction de leur taille d'effet. Parmi ces séquences, 72 ont été bénéfiques, 28 non bénéfiques et 14 ont été nuisibles. La figure 2.17 illustre l'écart entre les pourcentages attendus et ceux obtenus. Le pourcentage de séquences bénéfiques avec sa valeur de 63,2% est très en dessous de la valeur cible de 76%. Ce résultat est cohérent avec le résultat concernant l'hypothèse (H1). En effet, si les enseignants ne suivent pas les recommandations, il n'y a aucune raison de constater une croissance significative du taux de séquences bénéfiques par rapport à celui observé lors de l'étude exploratoire.

(H3) Le nouveau modèle d'orchestration reste valide pour l'enseignement secondaire Cet ensemble de résultats négatifs nous a permis de formuler une nouvelle

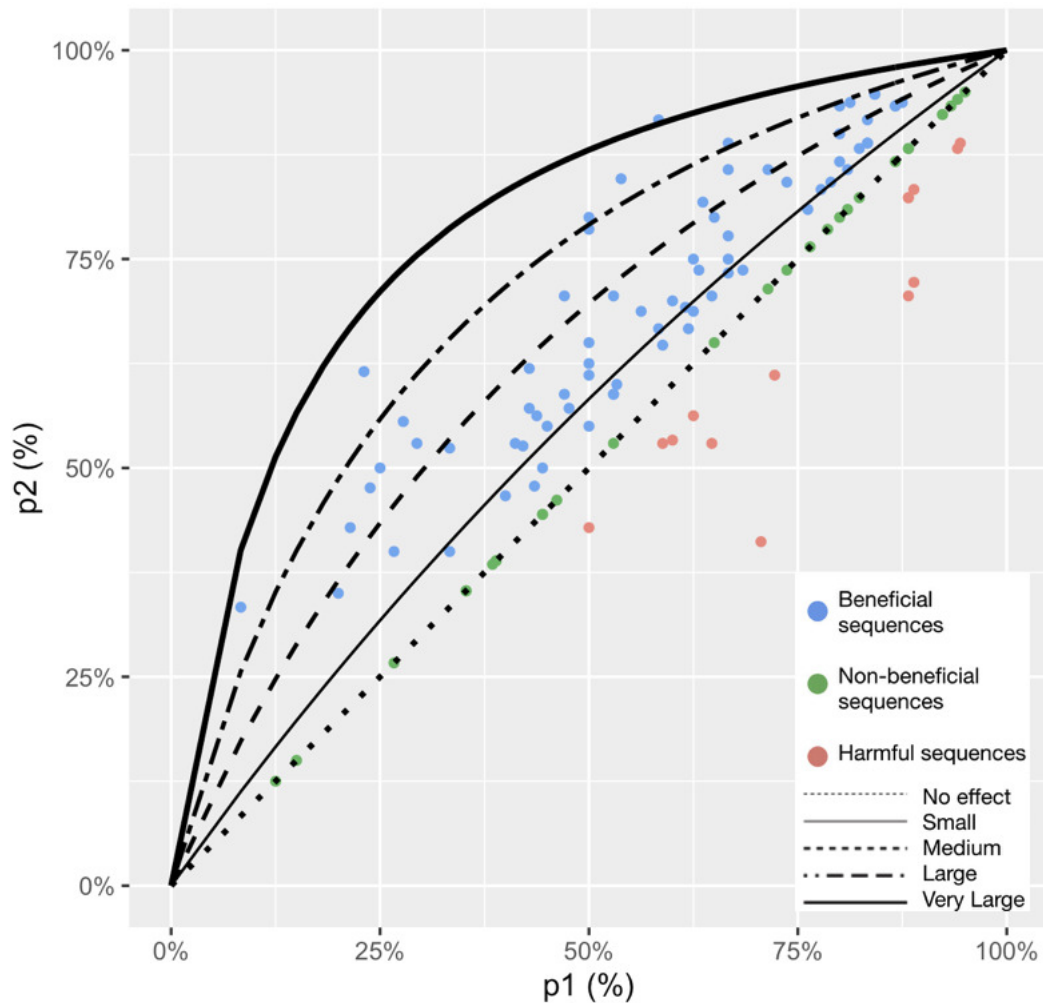


FIGURE 2.16 – Jeu de données 2022 - Séquences et lignes de la taille de l'effet de Cohen d : aucun effet ($d = 0$), petit ($d = 0,2$), moyen ($d = 0,5$), grand ($d = 0,8$) et très grand ($d = 1.2$).

hypothèse.

(H3) Le modèle d'orchestration proposé reste valide pour les séquences orchestrées dans l'enseignement secondaire.

Cette hypothèse est digne d'intérêt car, l'étude exploratoire ayant collecté des données sur l'utilisation d'Elaastic exclusivement dans l'enseignement supérieur, nous n'avons aucune indication sur la pertinence du modèle dans le secondaire. Le fait que les enseignants du secondaire n'aient pas suivi les recommandations nous a permis de mobiliser les tests statistiques utilisés dans le cadre de l'étude exploratoires sur le nouvel échantillon.

Les corrélations observées sur les variables caractérisant une séquence sont conformes à celles observées sur les données de l'étude exploratoire exceptée pour le lien entre la cohérence de l'évaluation par les pairs et la taille d'effet d'une séquence : les résultats du test sont non concluants sur cette relation.

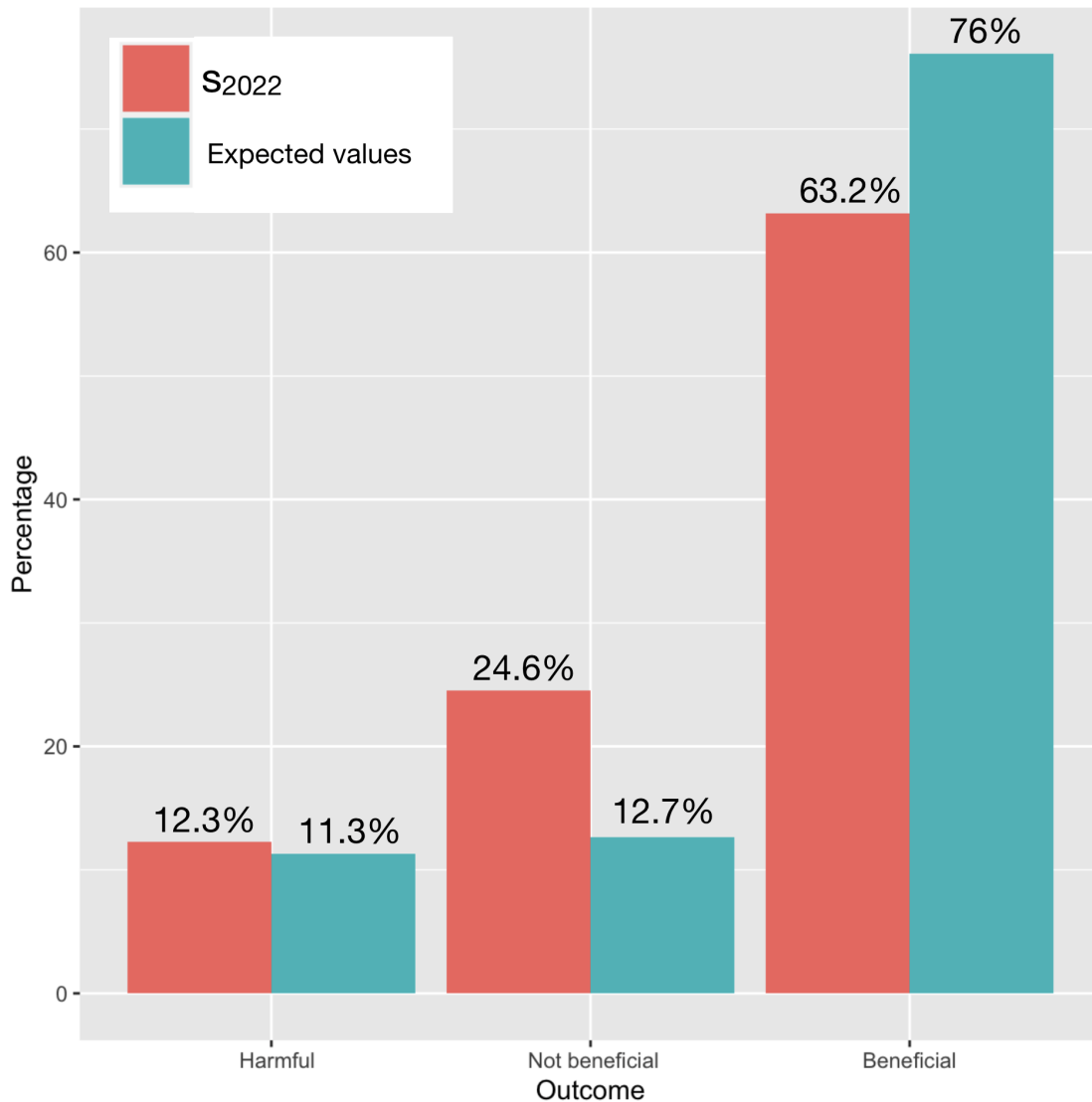


FIGURE 2.17 – Échantillon 2022 - Répartition des séquences comparée à la répartition attendue

2.4.6 Discussion et positionnement des travaux

[QR 2.1] Indicateurs prédictifs de la qualité d'une séquence

L'étude exploratoire a permis de mettre en lumière une série d'indicateurs ayant une influence sur la qualité d'une séquence d'évaluation formative se déroulant suivant un processus à 2 votes.

La proximité à 50% du pourcentage de bonnes réponses au premier vote La proximité à 50% du pourcentage de bonnes réponses au premier vote est prédictif du caractère bénéfique d'une séquence. Ce constat survient alors que l'algorithme, lors de la deuxième phase, propose aux apprenants, autant que faire ce peut, une répartition équilibrée des réponses à évaluer entre bonnes et mauvaises réponses. Ce qui signifie que, quel que soit le pourcentage réel de bonnes réponses au premier vote, les étudiants perçoivent

une répartition sensiblement équilibrée entre bonnes et mauvaises réponses. Malgré cela, dans le cas où le taux de bonnes réponses réel est éloigné de 50%, les chances d'obtenir une séquence bénéfique diminuent. En conséquence, les efforts pour limiter l'influence de la répartition réelle des votes n'ont pas d'impact quand le taux de réponses correctes au premier vote est trop faible ou trop élevé. Ce constat conforte le fait que le pourcentage de réponses correctes au premier vote est un indicateur objectif de la difficulté de la question pour une audience donnée.

La cohérence du degré de confiance À l'issue du premier vote, la cohérence du degré de confiance avec le caractère correct d'une réponse est un prédicteur de la qualité de la séquence d'évaluation par les pairs. En d'autres termes, si les apprenants ayant répondu correctement ne sont pas confiants alors que ceux qui se sont trompés sont confiants, il y a peu de chances d'obtenir des évaluations cohérentes : les bonnes réponses ont moins de chance d'être évaluées correctement alors que les mauvaises réponses peuvent recevoir de bonnes notes.

La cohérence de l'évaluation par les pairs À l'issue de la phase de confrontation de point de vue, lorsque les évaluations par les pairs sont incohérentes, les chances d'obtenir une séquence bénéfique diminuent significativement. Ce résultat est assez intuitif : si les réponses incorrectes sont mieux notées que les correctes, il y a peu de chance de pouvoir inverser la tendance de votes dans un sens positif.

L'autoévaluation dans le contexte de l'évaluation par les pairs Dans le processus implanté par Elaastic, un apprenant peut avoir à évaluer sa propre réponse durant l'activité d'évaluation par les pairs. Alors que la mise en situation d'autoévaluation est décrite comme bénéfique pour les apprentissages dans la littérature (MCMILLAN & HEARN, 2008 ; RATMININGSIH et al., 2017 ; ROLHEISER & ROSS, 2001), elle s'avère délétère dans le contexte d'une évaluation par les pairs telle qu'implantée dans Elaastic : en effet, les apprenants ont tendance à surnoter de manière significative leur propre réponse dans cette autoévaluation explicite, nuisant ainsi à la possibilité de changer d'avis quand ils se sont trompés (autoévaluation implicite). Des entretiens informels auprès de quelques étudiants permettent d'émettre des hypothèses expliquant le phénomène observé : les étudiants interrogés, ont indiqué qu'en se mettant une bonne note, ils avaient plus de chances de voir leur réponse prise en exemple pour la discussion lors de la phase de restitution des résultats. D'autre part, les étudiants ont mentionné la sensation d'incohérence à se donner une mauvaise note : « Je ne peux pas ne pas être d'accord avec moi-même ». On constate pourtant que les étudiants peuvent changer d'avis dans le cadre d'une séquence à 2 votes : ils peuvent donc « ne pas être d'accord avec eux-mêmes ». Ce constat pointe la différence d'impact métacognitif selon que la situation d'autoévaluation est explicite ou

implicite après avoir évalué d'autres réponses.

[QR 2.2-2.3] Recommandations auprès des enseignants et des concepteurs de système

Les recommandations en direction des enseignants ([QR 2.2]) induites par les indicateurs permettant de prédire la qualité d'une séquence sont décrites dans la section 2.4.3. Ces recommandations ont conduit à la création d'un nouveau modèle d'orchestration plus précis et plus complet que les modèles proposés précédemment sur les processus à 2 votes. Ce nouveau modèle est par ailleurs déterministe. L'enrichissement de ce nouveau modèle se fonde sur des indicateurs qui ne peuvent pas être calculés simplement par un humain en un temps raisonnable. En conséquence, ce modèle d'orchestration, pour être mis en œuvre, doit nécessairement être soutenu par la technologie. La recherche en IEIAH est particulièrement bien incarnée dans une telle situation, dans laquelle les bénéfices pédagogiques ne peuvent être atteints que si la technologie est utilisée en soutien. La question des lignes directrices et des recommandations de conception de système prend alors tout son sens. Ces recommandations à l'attention des concepteurs de systèmes de soutien aux processus à 2 votes ([QR 2.3]) sont présentées dans la section 2.4.3. Ces recommandations portent essentiellement sur la mise à disposition, dans une forme appréhendable par les enseignants, des indicateurs fondant le nouveau modèle d'orchestration. Ces indicateurs non triviaux induisent la prise en compte d'exigences d'explicabilité dans les systèmes à concevoir. Les résultats de l'expérimentation portant sur l'implantation du nouveau processus à 2 votes ont montré que l'exigence d'utilisabilité prévaut à l'exigence d'explicabilité : des recommandations non visibles, non actionnables par les utilisateurs, sapent littéralement les efforts menés pour satisfaire l'exigence d'explicabilité.

[QR 2] *Learning analytics* pour la compréhension et l'amélioration des processus à 2 votes

Les indicateurs prédictifs de la qualité d'une séquence d'évaluation formative et la proposition d'un nouveau modèle d'orchestration fondé sur ces indicateurs est un premier pas décisif dans la meilleure compréhension de ce qui est en jeu durant de telles séquences. Le nouveau modèle n'a cependant pas passé l'épreuve du terrain : les recommandations non visibles et donc non utilisables n'ont pas permis d'évaluer l'impact réel du nouveau modèle d'orchestration en contexte écologique.

BPMN or not BPMN

Durant ma thèse, j'avais pris le parti de représenter les processus d'évaluation d'audience en BPMN. Dans le cadre de la thèse de ANDRIAMISEZA, BPMN n'a pas été utilisé comme langage de représentation des processus à 2 votes. Le modèle de Vickrey sur lequel

s'appuie une partie du travail de Rialy n'a jamais fait l'objet de représentation en BPMN dans les publications le mentionnant. Le choix de conserver le formalisme simple utilisé dans les communications présentant les modèles d'orchestration de processus à 2 votes est un choix guidé par la volonté de faciliter la compréhension de l'évolution proposée du modèle. Nos travaux par ailleurs ne s'appuient pas sur BPMN pour implanter les processus dans nos solutions logicielles. À ce stade, la mobilisation de BPMN ne semble pas utile au développement de nos travaux.

Publications

Les travaux menés dans le cadre du projet B4MATIVE ont fait l'objet de plusieurs publications (ANDRIAMISEZA et al., 2021a; ANDRIAMISEZA et al., 2021b, 2021c) dont le récompensé *best paper* à EC-TEL 2021 (ANDRIAMISEZA et al., 2021b) ayant permis une publication d'un article étendu dans la revue *IEEE Transactions On Learning Technologies* (ANDRIAMISEZA et al., 2023). L'ensemble des travaux et l'étude portant sur l'analyse des données d'apprentissage auprès des enseignants du secondaire après l'implantation du nouveau modèle d'orchestration dans Elaastic sont consignés dans le manuscrit de thèse (ANDRIAMISEZA, 2022).

2.4.7 Rétrospective

Prise en compte de la rétrospective précédente

Capitalisation sur les travaux de recherche antérieurs La thèse de Rialy s'inscrit directement dans la continuité des travaux de ma thèse. L'ancrage théorique apparut tardivement dans la thèse précédente est mécaniquement présent au démarrage des travaux de Rialy.

Capitalisation sur Elaastic Elaastic est en production depuis 2015. Rialy commence sa thèse en 2019. Le socle technologique en place limite alors les développements à produire durant la thèse aux parties strictement utiles aux tests des hypothèses identifiées.

Intégration dans l'équipe de recherche d'une expertise en statistiques La collaboration avec Jean-François Parmentier se renforce après nos premières contributions communes (PARMENTIER & SILVESTRE, 2019; SILVESTRE & PARMENTIER, 2021). Son expertise en statistiques, sa connaissance encyclopédique des travaux relatifs aux pédagogies actives ainsi que sa culture d'approches expérimentales apportent une très précieuse valeur ajoutée aux travaux conduits durant la thèse de Rialy. Jean-François intègre l'équipe TALENT à sa création comme chercheur associé.

Augmentation significative des ressources humaines sur la thèse Les personnes investies sur la thèse furent le doctorant à temps plein, deux codirecteurs de thèse, un expert en statistiques, et un ingénieur de recherche pour l’encadrement technique. La thèse ne devient plus le travail d’un doctorant et de son ou ses encadrants : la thèse devient un travail d’équipe intégrant des compétences variées afin d’obtenir les meilleurs résultats possibles.

Un environnement pour les expérimentations extérieur à l’environnement de recherche La thèse se déroule en partenariat avec l’Académie de Nancy-Metz dont la mission principale est la fourniture des terrains d’expérimentations.

Les forces

La gestion des risques La thèse de Rialy a été traversée par la crise sanitaire relative au COVID 19. Une approche dirigée par la gestion des risques a permis d’orienter le travail du doctorant très tôt (avant que le confinement soit effectif en France) sur l’analyse des données existantes collectées entre 2015 et 2019. Nous anticipions ainsi le risque lié à l’impossibilité de mener les expérimentations dans les établissements de l’Académie Nancy-Metz.

Une étude exploratoire choisie Alors que l’approche exploratoire durant ma thèse relevait davantage du tâtonnement que d’une démarche choisie, l’étude exploratoire de la thèse de Rialy s’intègre dans un plan de thèse structuré, guidé par une gestion de risques.

Introduction de l’approche *Worked example* pour la rédaction d’article Après ma thèse, l’intérêt porté à la théorie de la charge cognitive (SWELLER, 2011) me conduit à étudier le cadre de conception pédagogique 4C/ID (VAN MERRIËNBOER & KIRSCHNER, 2017). Ce cadre attache une grande importance à la manière de réguler le guidage et le support proposés aux apprenants durant leurs activités d’apprentissage. 4C/ID mobilise ainsi l’approche par étude d’exemples résolus (ATKINSON et al., 2000 ; SWELLER, 2006) pour accompagner l’apprentissage des novices sur le développement de nouvelles compétences. Durant la thèse de Rialy, nous avons utilisé l’approche *worked examples* dans le cadre de la formation doctorale en recherchant des articles pouvant servir d’exemples à la publication de nos travaux. La structure de l’article de GUO et al. (2014) a directement inspiré la structure de l’article EC-TEL 2021 ayant reçu le prix du meilleur papier (ANDRIAMISEZA et al., 2021b).

Un plan de publication exemplaire L’obtention du prix du meilleur papier à EC-TEL 2021 a ouvert la porte de la publication d’un article dans la revue *IEEE Transactions*

on *Learning Technologies*¹⁰ de rang Q1 dans les champs *Computer science Applications, Education et E-learning*¹¹. Les travaux menés dans la cadre de la thèse de Rialy ont donc fait l'objet de publications dans une conférence francophone, puis dans des conférences internationales reconnues dans le domaine jusqu'à une dernière publication en revue internationale de premier rang. Nous aspirons à reproduire ce type de trajectoire de publications dans nos encadrements de thèses en cours.

Les faiblesses

Peu de place pour les expérimentations contrôlées L'implantation du modèle d'orchestration proposé par Rialy dans Elaastic a permis de collecter des données d'utilisation en contexte écologique. Cependant, les utilisations par les enseignants n'ont pas fait l'objet d'un plan qui aurait sans doute permis de contrôler davantage les variables que nous souhaitions étudier.

Des choix techniques et stratégiques questionnables Lorsque que nous avons élaboré le projet de recherche avec l'Académie de Nancy-Metz, celle-ci était en train de déployer Moodle sur son territoire. Le choix de donner accès aux enseignants à Elaastic via Moodle résolvait un problème technique (nous pouvions gérer la création de compte et l'authentification via le protocole LTI déjà supporté par Elaastic) et un problème stratégique : Elaastic faciliterait l'adoption de Moodle en proposant aux enseignants un service pédagogique qui devait les intéresser fortement. Ses choix ne menèrent finalement pas aux résultats attendus. L'utilisation de Moodle pour utiliser Elaastic apparut comme un frein aux usages. En effet, tous les efforts de formation et d'accompagnement devaient porter d'abord sur l'appropriation des rudiments de Moodle avant de concerner l'utilisation d'Elaastic. Nous n'avons pas mesuré précisément l'impact de cette approche sur le faible taux d'utilisation d'Elaastic par rapport au nombre de formations et au temps passé en accompagnement, mais il est certain que cela a augmenté les difficultés déjà présentes en période de crise sanitaire liée au COVID 19.

2.5 Conclusion

Nous avons présenté dans ce chapitre les travaux en lien avec la conception de systèmes technologiques soutenant la mise en œuvre de séquences d'évaluation formative en classe ou en amphithéâtre.

La première partie décrit les travaux menés durant ma thèse conduisant à la proposition d'un processus d'évaluation d'audience et de son implantation associée au sein de la plateforme Tsaap-Notes. Le processus en N Phases vise à couvrir au mieux les activités et

10. <https://ieeexplore.ieee.org/xpl/RecentIssue.jsp?punumber=4620076>

11. <https://www.scimagojr.com/journalsearch.php?q=19700167026&tip=sid>

stratégies caractéristiques de l'évaluation formative. La plateforme Tsaap-Notes soutient les enseignants et les apprenants dans la mise en œuvre du processus.

La deuxième partie du chapitre se focalise sur les travaux relatifs à la thèse de Rialy Andriamiseza que j'ai codirigée avec Julien Broisin. Ces travaux s'appuient sur une analyse exploratoire des données collectées durant l'utilisation de la plateforme Elaastic entre 2015 et 2019. Elaastic est une évolution de Tsaap-Notes implantant un processus dérivé du processus en N Phases. Cette analyse met en lumière des indicateurs permettant de prédire la qualité d'une séquence d'évaluation formative dans le contexte d'un processus à 2 votes. De ces indicateurs ont émergé des recommandations pour les enseignants et concepteurs de système afin d'améliorer la qualité des séquences d'évaluation formative.

Le chapitre suivant présente le deuxième axe de recherche initié durant ma thèse et développé durant les 5 dernières années.

Chapitre 3

La technologie en soutien de l'évaluation formative hors classe

Contenu

3.1	État de l'art	52
3.1.1	Les tests autocorrectifs	52
3.1.2	Les effets test et espacement	53
3.2	Problématiques et questions de recherche	53
3.2.1	L'automatisation de la construction de tests autocorrectifs . . .	53
3.2.2	La conception d'un système de révision	55
3.3	La génération semi-automatique de tests autocorrectifs . . .	55
3.3.1	Méthode de conduite de la recherche	55
3.3.2	Première itération	56
3.3.2.1	L'approche <i>Note as feedback</i>	56
3.3.2.2	Expérimentation et résultats	57
3.3.3	Seconde itération	60
3.3.3.1	L'approche <i>Rationale as feedback</i>	60
3.3.3.2	Expérimentation et résultats	61
3.3.4	Discussion et positionnement des travaux	62
3.4	La conception d'un système de révision	64
3.4.1	Méthode de conduite de la recherche	64
3.4.2	L'étude qualitative exploratoire	65
3.4.2.1	Participants et collecte des données	66
3.4.2.2	Résultats et hypothèses induites	67
3.4.2.3	L'application Konsolidation	69
3.4.3	Discussion et positionnement des travaux	69
3.4.4	Rétrospective	71
3.5	Conclusion	72

3.1 État de l'art

Une partie du travail conduit durant ma thèse se focalisait sur la mise en œuvre d'évaluations formatives régulées et orchestrées par l'enseignant dans des conditions d'enseignement en face à face. Les tests autocorrectifs permettent la mise en œuvre d'évaluations formatives se déroulant hors la classe, en l'absence d'enseignant.

3.1.1 Les tests autocorrectifs

Dès 1993, plusieurs travaux (CLARIANA, 1993 ; ZAKRZEWSKI & BULL, 1998) montrent l'efficacité de dispositifs d'auto-évaluation informatisés pour la mise en œuvre d'évaluations formatives dans l'enseignement supérieur.

Différents travaux (MILLER, 2009 ; RICKETTS & WILKS, 2002) montrent que les évaluations informatisées facilitent la mise en œuvre d'évaluations formatives fréquentes pour des cohortes comprenant un grand nombre d'étudiants. La fourniture d'un *feedback* aux étudiants relatif à leurs travaux ou à leurs résultats d'évaluation est une clé de succès pour l'amélioration de leur apprentissage (HATTIE & TIMPERLEY, 2007 ; HIGGINS, 2000b). NICOL et MACFARLANE-DICK préconisent ainsi l'utilisation de tests en ligne pour que le feedback puisse être mis à disposition des apprenants à n'importe quel instant, depuis n'importe quel lieu, aussi souvent que l'apprenant le souhaite (NICOL & MACFARLANE-DICK, 2006).

La mise en place des tests autocorrectifs est triviale pour évaluer la maîtrise de connaissances (utilisation de questionnaires mobilisant des questions à choix multiples, des questions d'appariement, des textes à trous, etc.). La compréhension de concepts peut aussi être évaluée avec des questions fermées : les travaux sur les *conceptests*, correspondant à un lot de questions permettant de tester la compréhension d'une notion plutôt que la mémorisation superficielle d'une formule ou d'un énoncé, rendent possible la correction automatique de questions fermées évaluant la compréhension des apprenants sur des concepts dans des domaines variés (FALCONER, 2004 ; MCCONNELL et al., 2006 ; SCHLATTER, 2002). L'accès aux questionnaires interactifs des études PISA permet de consulter un grand nombre de questions fermées, qui peuvent donc être corrigées de manière automatique, visant à évaluer des compétences de résolution de problèmes en mathématiques et en sciences (OCDE, 2022).

La pertinence des tests autocorrectifs pour améliorer les résultats d'apprentissage est par ailleurs soutenue par les résultats issus de la recherche en psychologie cognitive ayant mis en lumière les effets test et d'espacement.

3.1.2 Les effets test et espacement

La manière de travailler en dehors des cours, dans le contexte de révisions pour un contrôle par exemple, est un critère déterminant dans l'amélioration des résultats. Se tester sur de nouvelles connaissances ou de nouveaux concepts est plus efficace que de relire le cours sur ces nouvelles notions. Ce phénomène est connu sous l'appellation d'effet test ou *testing effect* en anglais (PHELPS, 2012 ; ROWLAND, 2014). Plusieurs études ont montré l'impact positif de l'effet test sur la mémorisation à long terme (ROEDIGER & KARPICKE, 2006 ; ZAROMB & ROEDIGER, 2010).

Lorsque les tests sont proposés à différentes occasions espacées dans le temps, l'effet produit sur l'apprentissage est plus important (CARPENTER et al., 2012 ; PHELPS, 2012 ; ROHRER & TAYLOR, 2006). Ce dernier constat a été confirmé par plusieurs études montrant que le travail réparti en plusieurs petites séquences étalées dans le temps est plus efficace en termes d'apprentissage que le travail concentré sur quelques séquences intensives. Cet effet est connu sous le nom d'effet d'espacement ou *spacing effect* en anglais (CEPEDA et al., 2006 ; DEMPSTER & FARRIS, 1990). Bien que l'essentiel des travaux portent sur l'acquisition et la mémorisation de connaissances relativement simples portant sur des informations verbales (CARPENTER et al., 2012), des études ont montré les bénéfices de l'effet d'espacement sur des tâches d'apprentissage jugées plus complexes dans différents domaines : entre autres exemples, l'étude de BIRD (2011) montre les bénéfices de l'effet sur des apprentissages mobilisant la compréhension de règles de grammaire alors que ROHRER et TAYLOR (2006) ont montré les bénéfices de l'approche pour la résolution de tâches mathématiques modérément abstraites.

LATIMIER (2019) dans sa thèse a étudié l'articulation des effets test et d'espacement, et souligné que leurs effets sont additionnels. Ainsi il apparaît pertinent d'appliquer conjointement ces deux effets au sein d'un même outil, pour espérer maximiser les résultats des apprentissages.

3.2 Problématiques et questions de recherche

3.2.1 L'automatisation de la construction de tests autocorrectifs

Les tests autocorrectifs, en tant que dispositif d'évaluation formative se prolongeant en dehors des cours, apparaissent comme un outil pertinent pour les activités de révisions des apprenants. Ces tests, pour améliorer leur efficacité, doivent fournir un feedback à chaque apprenant répondant aux différentes questions du test.

En raison de l'augmentation croissante du nombre d'étudiants en Université, la vision de l'enseignement consistant à rendre l'enseignant seul source du feedback à adresser à chaque étudiant n'est pas soutenable (NICOL & MACFARLANE-DICK, 2006). Les expériences décrites par NICOL et BOYLE (2003) mettent en valeur l'usage des systèmes de

votes interactifs durant les cours en face à face pour la constitution de feedbacks provenant à la fois de l'enseignant et des étudiants. Cette exploitation des productions des étudiants comme feedback est explorée plus en avant dans les travaux présentés dans (LEA, 2001) : les messages postés par les étudiants dans des forums sont archivés par l'enseignant pour être réutilisés ultérieurement comme feedback adressé à d'autres étudiants.

Dans les systèmes généralistes fournis par les plates-formes d'enseignement en ligne (Moodle, Chamilo, Claroline, Blackboard ou autres dispositifs compatibles avec la spécification IMS QTI), le feedback proposé à l'étudiant est par défaut minimaliste : il comprend, pour chaque question, le score de l'étudiant ainsi que la réponse correcte. Si le concepteur de la question a pris la peine de compléter les caractéristiques détaillées de la question lors de son édition avec des éléments complémentaires (tels que des explications générales, des explications pour chaque réponse, ou des ressources additionnelles), alors le système est capable de proposer ce feedback complémentaire à l'étudiant. Une telle gestion de ces informations pose cependant différents problèmes : (1) la saisie par l'enseignant du feedback complémentaire pour chaque question est consommatrice de temps si bien qu'elle n'est pas toujours réalisée, (2) le feedback proposé par le système est identique pour l'ensemble des étudiants (BULL & MCKENNA, 2003), (3) le feedback saisi par l'enseignant n'est pas toujours compris par les étudiants, car il n'est pas rédigé dans leur « dialecte » (CHANOCK, 2000 ; HIGGINS, 2000a), et enfin (4) le feedback n'est pas forcément mis en relation avec la structure et le déroulement du cours alors que cette caractéristique représente une des garanties de l'efficacité du feedback (BULL & MCKENNA, 2003).

Ces constats nous ont conduits à la question de recherche suivante :

[QR 3] Comment automatiser, autant que faire se peut, la construction de tests de révisions afin d'alléger la charge de travail des concepteurs de tests ?

Plus spécifiquement, nous nous sommes intéressés à la fourniture de feedback textuel dans ces tests, ce qui nous a amené à traiter la question suivante :

[QR 3.1] Comment générer automatiquement du feedback textuel dans les tests de révisions autocorrectifs ?

Ces questions ont été abordées durant ma thèse et ont suscité un fort intérêt auprès de la communauté des praticiens et des chercheurs se traduisant, entre autres, par l'obtention de la distinction *best paper* à la conférence TICE 2014 (SILVESTRE et al., 2014a), et différentes publications dans les conférences internationales ICALT et EC-TEL (SILVESTRE et al., 2015a, 2015c). Ces résultats encourageants nous ont conduits à poursuivre des travaux dans cette direction, en nous intéressant au champ plus large de la conception d'un système de révisions.

3.2.2 La conception d'un système de révision

Le premier principe ayant guidé nos travaux sur la conception d'un système de révision est le suivant : le système conçu doit soutenir les effets test et d'espacement dont les impacts positifs sur les apprentissages ont été démontrés.

Les effets test et d'espacement sont mis en œuvre dans de nombreuses applications numériques tels que Duolingo, Babbel, Mosalingua, Anki, etc. Ces applications sont conçues essentiellement pour l'apprentissage des langues ou de connaissances simples (LATIMIER, 2019; SEIBERT HANSON & BROWN, 2020). Elles n'ont pas été conçues pour une utilisation dans un cadre scolaire et n'intègrent donc ni les enseignants, ni les autres travaux à réaliser par les élèves hors la classe dans les activités proposées. LATIMIER souligne le manque d'études sur les effets test et espacement conduites en contexte écologique (scolaire ou universitaire) ainsi que l'absence d'étude sur l'efficacité des effets en fonction des différents niveaux d'études. Dans le même temps, LATIMIER (2019) dont l'étude réalisée en situation expérimentale s'est appuyée sur l'utilisation de la plateforme numérique Didask, rappelle l'intérêt de mobiliser les outils numériques pour l'investigation des processus d'apprentissage et de leur optimisation. Ces constats nous conduisent à formuler un second principe de conception : le système conçu doit s'intégrer dans le contexte écologique d'un enseignement institutionnalisé, scolaire ou universitaire.

La question de recherche émergeant des deux principes de conception peut se formuler comme suit :

[QR 4] Comment concevoir un système de révisions utilisable en contexte scolaire ou universitaire soutenant les effets test et d'espacement ?

Ces travaux de recherche sont l'objet de la thèse en cours de Denis Ollivier.

3.3 La génération semi-automatique de tests auto-correctifs

3.3.1 Méthode de conduite de la recherche

Le travail proposé s'est inscrit dans une démarche itérative et incrémentale très influencée par notre expérience dans le développement logiciel et plus spécifiquement dans la mise en œuvre des méthodes agiles et du *Lean Startup* (RIES, 2012). Nous n'étions pas en 2014 informés des travaux sur le *Design Based Research* ce qui explique certains écarts par rapport au DBR : d'une part, la méthode mise en œuvre n'a impliqué ni enseignant, ni étudiant dans la définition des problèmes traités et, d'autre part, les résultats

des travaux n'ont pas pris la forme de recommandations ou de formulations de principes de conceptions.

Deux itérations ont structuré le travail. Toutes deux ont eu pour objectif de résoudre des problèmes concrets rencontrés en contexte écologique à l'aide de la fourniture d'un système technologique. Des expérimentations en contexte écologique ont permis d'éprouver les systèmes conçus à l'issue de chaque itération. Les limites mises en lumière à l'issue de la première itération ont guidé les travaux de la seconde.

3.3.2 Première itération

3.3.2.1 L'approche *Note as feedback*

Les travaux menés sur la conception de systèmes d'évaluation formative (cf. chapitre 2) conduisent à la constitution, pour chaque enseignement mobilisant le système Tsaap-Notes ou Elaastic, d'une base de questions directement en lien avec le cours dispensé par l'enseignant. Notre première idée a donc été de générer des tests de révisions en recyclant les questions posées pendant les cours. Moodle étant une des plateformes LMS les plus populaires dans les universités du monde entier¹, nous nous sommes concentrés sur la génération automatique d'un fichier constituant un test dans un format compatible avec la plateforme Moodle.

Afin d'intégrer dans ces tests de révisions des feedbacks textuels tout en préservant les enseignants de la construction de ces feedbacks, nous avons recyclé les productions textuelles des étudiants issues des séquences d'évaluation formative orchestrées avec Tsaap-Notes.

La première version de Tsaap-notes a été conçue comme un dispositif combinant un système de vote interactif et un dispositif de prise de notes collaborative (SILVESTRE et al., 2014a, 2015a). L'objectif était alors d'encourager (1) les échanges constructifs entre enseignant et étudiants suite à la présentation des résultats obtenus aux différentes questions, et (2) l'appropriation de ce feedback par les étudiants à travers la prise de notes. La figure 3.1 présente les résultats d'une question posée durant un cours portant sur le développement d'applications distribuées, accompagnée d'une note prise par un étudiant au cours de la restitution des résultats.

Lorsqu'un enseignant génère le fichier de test de révisions à partir de Tsaap-Notes (cf. figure 3.2), les notes prises par les étudiants au cours des échanges suivant la restitution des résultats d'une question sont intégrées comme feedback textuel associé cette question.

Une fois le fichier généré par Tsaap-Notes récupéré, l'enseignant peut l'importer dans Moodle. Les questions peuvent alors être éditées via l'interface du LMS et utilisées comme questions pour constituer un test de révisions Moodle. La figure 3.3 illustre l'édition d'une question Tsaap-Notes importée dans Moodle : les notes prises par les étudiants spécifient

1. <https://stats.moodle.org/>



FIGURE 3.1 – *Note as feedback* : affichage des résultats d'une question avec note prise par un étudiant.

le feedback général textuel associé à la question. L'enseignant peut alors modifier ou supprimer certaines de ces notes.

Lorsqu'un étudiant a répondu à une question d'un test Moodle conçu à partir de Tsaap-Notes, les notes prises par l'ensemble des étudiants sont présentées lors de la restitution des résultats. La figure 3.4 illustre l'affichage du feedback textuel dans le contexte de la restitution des résultats à une question.

3.3.2.2 Expérimentation et résultats

L'expérimentation visait la mise à l'épreuve des hypothèses suivantes :

- (H1) Le recyclage des notes des étudiants dans les tests de révisions engage davantage les étudiants dans l'activité de prise de notes sur les questions interactives.
- (H2) Les étudiants participent massivement aux tests de révisions.
- (H3) les étudiants, mieux préparés, obtiennent de meilleurs résultats aux évaluations sommatives.

Le dispositif *Note as Feedback* a été expérimenté en 2014 sur un groupe de 54 étudiants inscrits en 1ère année de Master Informatique dans le cadre de l'enseignement intitulé « Développement d'applications réparties avec Java EE ». Sur les 54 étudiants inscrits, 35, en moyenne, furent présents pendant les 10 heures de cours dispensées en face-à-face

The screenshot shows the Tsaap-Notes interface. At the top, there is a navigation bar with 'Tsaap-Notes', 'Notes', and 'Scopes' on the left, and 'Franck Silvestre' with a settings gear icon on the right. Below this is a breadcrumb trail: 'Scope List / Show scope "Master/Informatique/DL/2013_2014/JavaEE/GORM-avance"'. A table displays details for this scope:

Scope Name	Master/Informatique/DL/2013_2014/JavaEE/GORM-avance
Url	http://moodle.univ-tlse3.fr/mod/resource/view.php?id=66510
Description	#gorm #hibernate #pagination #n+1select #locking
Owner	@FranckSilvestre (as teacher)
Created	2014-03-25 03:21:55 CET
Updated	2014-03-25 03:21:55 CET
The notes on this scope	https://notes.tsaap.eu/tsaap-notes/notes/index?displaysAll=on&contextName=Master%2FInformatique%2FDL%2F2013_2014%2FJavaEE%2FGORM-avance&contextId=18

Below the table, there are two buttons: 'Edit' and 'Export questions'. The 'Export questions' button is circled in red. At the bottom, there is a copyright notice: '© Tsaap Development Group 2013 - Tsaap-Notes version 2.1.1 - Mentions'.

FIGURE 3.2 – *Note as feedback* : la fonctionnalité de génération du fichier au format Moodle.

dans le cadre de cet enseignement, soit un taux de présence moyen de 65%. D'autre part, au total, 82 questions interactives ont été posées aux étudiants via la plate-forme Tsaap-Notes. Le nombre moyen d'étudiants répondant à ces questions s'élève à 26, soit un taux moyen de participation aux questions interactives égal à 74%.

Participation à la prise de notes sur les questions interactives Au démarrage de l'expérimentation (durant la première heure de cours), la plate-forme Tsaap-Notes a été rapidement présentée aux étudiants. D'autre part, nous leur avons indiqué que les questions interactives seraient injectées dans Moodle sous la forme de tests de révision. Durant les 6 premières heures de cours, l'exportation des questions avec recyclage des notes prises par les étudiants n'était pas encore une fonctionnalité disponible au sein de Tsaap-Notes ; les étudiants n'ont donc pas été informés que les notes seraient réutilisées dans les tests de révision. Sur cette première partie d'expérimentation, très peu de notes furent prises sur les résultats des questions interactives : pour 55 questions posées, seulement 6 notes furent postées sur la plate-forme, soit un taux de 10%. Au démarrage de la 7^e heure de cours, les étudiants ont été informés de la réutilisation de leurs notes dans les tests de révision. Sur les 4 dernières heures de cours, pour 27 questions interactives posées, 14 notes ont été postées par les étudiants, soit un taux de 52%.

3.3. LA GÉNÉRATION SEMI-AUTOMATIQUE DE TESTS AUTOCORRECTIFS 59

Catégorie actuelle Import tsaap-Notes (19) Utiliser cette catégorie

Enregistrer dans la catégorie Import tsaap-Notes (19)

Nom de question* Gorm N+1 select

Texte de la question

Police Taille police Paragraphe

B *I* U ABC X₂ X₁ [Liste à puces] [Liste ordonnée] [Alignement] [Couleur] [Style] [HTML]

En supposant que la BDD contienne 17 acteurs Combien de requêtes SQL génère le code suivant par défaut "Acteur.list(offset:10, max 10).each { println it.role }"

Chemin: p

Note par défaut* 1

Feedback général

Police Taille police Paragraphe

B *I* U ABC X₂ X₁ [Liste à puces] [Liste ordonnée] [Alignement] [Couleur] [Style] [HTML]

Feedback coming from notes taken in Tsaap-Notes...

@EvansBis: Il recupère d'abord les acteur puis effectue une requête sur chacun d'entre eux, à savoir : 1 + 7

@LoicV: 8 requêtes -> 1 requête pour la liste d'acteur et 7 requêtes pour les rôles

Chemin: p

FIGURE 3.3 – *Note as feedback* : édition d'une question issue de Tsaap-Notes et importée dans Moodle.

Question 15
Incorrect
Note de 0,00 sur 1,00
Modifier la question

En supposant que la BDD contienne 17 acteurs Combien de requêtes SQL génère le code suivant par défaut "Acteur.list(offset:10, max 10).each { println it.role }"

Veillez choisir une réponse :

a. 17 requêtes

b. 7 requêtes **X**

c. 8 requêtes

d. 10 requêtes

e. 1 requête

Feedback coming from notes taken in Tsaap-Notes...

@EvansBis: Il recupère d'abord les acteur puis effectue une requête sur chacun d'entre eux, à savoir : 1 + 7

@LoicV: 8 requêtes -> 1 requête pour la liste d'acteur et 7 requêtes pour les rôles

La réponse correcte est : 8 requêtes

FIGURE 3.4 – *Note as feedback* : affichage du feedback dans Moodle pour une réponse donnée par un étudiant participant à un test de révisions.

Le premier résultat notable de cette expérimentation est que la réutilisation des notes prises par les étudiants dans les tests de révision est l'explication la plus probable du renforcement de l'activité de prise de notes sur les questions interactives : l'utilité concrète des notes a incité à la prise de notes.

Participation des étudiants aux tests de révision Sur le cours Moodle correspondant à l'enseignement faisant l'objet de notre expérimentation, deux séries de tests ont été créées (la première ayant été mise à disposition des étudiants avant la seconde) : une série ne réutilise pas les notes formulées par les étudiants, alors que l'autre s'appuie sur ces notes comme source de *feedback*. Les tests sans recyclage des notes ont été ignorés par la quasi-totalité des étudiants : 4 étudiants ont démarré les tests dépourvus de *feedback*, parmi lesquels 3 ont basculé sur les tests intégrant des *feedback* dès leur mise à disposition. 47 étudiants, soit 87% des inscrits, ont réalisé les tests de révisions avec *feedback*.

Apparaît ici un deuxième résultat notable : les étudiants préfèrent à l'unanimité jouer les tests avec feedbacks issus de leur prise de notes, plutôt que les tests sans feedback. Nous pouvons interpréter ce résultat comme le fait que les étudiants considèrent les notes prises pendant le cours comme aide durant les tests de révision.

Résultats au test d'évaluation sommative À la fin de la période de cours et de révisions, un test Moodle comprenant 102 questions et comptant pour 30% de la note finale de l'examen a été soumis aux étudiants. Parmi les 102 questions, les 82 questions issues de Tsaap-Notes ont été réutilisées, mais les valeurs numériques des questions ont été modifiées. Les résultats obtenus cette année (2014) ont été radicalement différents de ceux constatés l'année précédente (le cours est en place depuis deux ans uniquement dans son contenu actuel) : la moyenne obtenue au test est de 17/20 avec un écart-type de 1,48 alors que la moyenne obtenue l'année précédente (2013) était de 11,92 avec un écart-type de 3,26. Cette comparaison avec les résultats de l'année précédente est cependant à prendre en compte avec modération en raison des différences notables entre les deux évaluations sommatives : en 2013 le test était composé de 54 questions contre 102 en 2014 ; seules 20 questions ont été posées à la fois en 2013 et 2014.

Le dispositif employé dans le cadre de notre enseignement mène au constat d'une hausse remarquable des résultats des étudiants sur la partie d'examen informatisée par rapport à l'année précédente. Il est cependant impossible de discerner ce qui, dans ce dispositif, a eu l'influence la plus notable : en 2013, la plate-forme Tsaap-Notes n'était pas intégrée à l'enseignement expérimenté ici et les tests d'autoévaluation n'avaient pas été proposés aux étudiants.

3.3.3 Seconde itération

3.3.3.1 L'approche *Rationale as feedback*

Malgré les résultats encourageants obtenus lors de la première expérimentation, deux limites ont clairement été identifiées dans le dispositif « Note as feedback » :

- (1) bien que la prise de notes augmente quantitativement lorsque les étudiants savent que leurs contributions seront la seule source de feedback dans les tests de révision,

relativement peu d'étudiants participent à l'activité de rédaction des explications nécessaires à la constitution du feedback (moins d'une note prise par question en moyenne) ;

- (2) toutes les notes saisies par les étudiants sur une question donnée sont récupérées de manière automatique pour constituer le feedback du futur test de révision, elles ne sont ni évaluées ni filtrées. Il n'y a donc pas de « contrôle qualité » du feedback intégré aux tests de révision.

La conception du processus en N Phases présenté en détail dans le chapitre 2 a été initialement motivée par la levée des limites constatées dans le dispositif *Note as feedback*. Nous baptisons *Rationale as feedback* le dispositif résultant de la mise en œuvre du processus en N Phase pour constituer les feedbacks dans les tests de révisions.

Pour augmenter le taux de participation à la rédaction d'explications sur une question posée pendant le cours, nous avons émis l'hypothèse que la tâche de rédaction des explications devait être insérée dans une activité fournissant un feedback immédiat à l'étudiant sur sa production. Ainsi, plutôt que d'espérer une prise de notes par quelques étudiants restituant le feedback de l'enseignant lors de la restitution des résultats sur une question, le processus en N Phases demande à chaque étudiant une explication justifiant son choix et lui propose un feedback immédiat sur sa contribution textuelle. Ce feedback prend plusieurs formes, allant de la confrontation à des réponses alternatives jusqu'à la présentation du résultat de l'évaluation par les pairs de sa contribution.

Une évaluation rapide d'un grand nombre d'explications ne peut pas être prise en charge par l'enseignant pendant le cours. Pour fournir rapidement à chaque étudiant une évaluation de son travail, nous avons opté pour l'évaluation par les pairs dont les vertus ont été mises en valeur dans différentes études (DOCHY et al., 1999 ; FREEMAN & MCKENZIE, 2014 ; K. TOPPING, 1998 ; K. J. TOPPING, 2009).

La deuxième limite révélée par les expérimentations du dispositif *Note as feedback* est l'absence de filtrage dans la sélection des contributions des étudiants pour la constitution des feedbacks dans les tests de révision générés par Tsaap-Notes. L'évaluation des explications par les étudiants permet la mise en œuvre d'un algorithme ne sélectionnant que les contributions les mieux notées par les étudiants. La génération automatique du fichier d'export Moodle contenant les questions et feedbacks du test de révision a été modifiée afin que les feedbacks ne soient constitués, pour une question donnée, que des explications ayant une évaluation moyenne supérieure à 2,5/5. La figure 3.5 illustre la nouvelle version du feedback dans Moodle produit à partir des réponses textuelles des étudiants.

3.3.3.2 Expérimentation et résultats

L'expérimentation menée sur le dispositif *Rationale as feedback* est l'expérimentation portant sur le processus en N Phases décrite en détail dans le chapitre 2. Nous nous

Soit L une librairie javascript distribuée sous licence GPL . Soit E une entreprise qui modifie la librairie L pour en faire une version L2. L'entreprise E propose à des utilisateurs son application serveur qui utilise la version L2 de la librairie. Est ce que les utilisateurs peuvent exiger de récupérer le code source de la version L2 du logiciel ?

Veillez choisir une réponse :

a. Oui

b. Non ✗

Feedback coming from notes taken in Tsaap-Notes...

3.88 /5 @superpedro: Le javascript étant exécuté côté client, l'utilisateur a une copie, la contamination du GPL force donc la possibilité d'accès au code à l'utilisateur.

2.67 /5 @Linkval: Parce que javascript est visible via navigateur. Magique

2.67 /5 @JeSuisCharlie: Le code source est visible par les utilisateurs et donc ils peuvent l'exiger.

2.62 /5 @Alex: L2 est en licence GPL, et s'exécute sur le poste de l'utilisateur. Il est donc titulaire de la copie, et peut donc demander le code source.

La réponse correcte est : Oui

FIGURE 3.5 – *Rationale as feedback* : affichage du feedback dans Moodle pour une réponse donnée par un étudiant participant à un test de révision. Seules les explications ayant une évaluation moyenne supérieure à 2,5/5 sont affichées.

focalisons ici sur un résultat déjà présenté, en l'abordant sous l'angle spécifique de la levée d'une des limites du dispositif *Note as feedback*.

Pour rappel, une des hypothèses testées par l'expérimentation était la suivante :

(H1) Le processus engage fortement les étudiants sur l'activité de rédaction d'explications textuelles.

Nous avons alors indiqué le résultat suivant : 68,3% des étudiants ont fourni une réponse textuelle en complément de la sélection d'un item. Cela signifie que pour une question posée, nous disposons dans le dispositif *Rationale as feedback* de plusieurs dizaines d'explications textuelles candidates à la fonction de feedback dans les tests de révisions, contre moins d'une en moyenne dans le dispositif *Note as feedback*. Cette manne d'explications candidates peut alors être triée puis filtrée en fonction de la valeur obtenue aux évaluations par les pairs.

3.3.4 Discussion et positionnement des travaux

[QR 3] L'automatisation autant que faire se peut de la construction de tests de révisions Les dispositifs *Note/Rationale as feedback* mis en œuvre pour faciliter la génération des tests de révisions n'ont automatisé que partiellement la construction de ces tests : en se limitant à la génération automatique du fichier contenant les questions d'un test, l'enseignant doit encore manuellement importer les questions dans Moodle et constituer le test à partir des questions importées.

Il n'y a pas réellement de freins techniques à l'automatisation complète de la construction de tests de révisions. Que ce soit en maintenant la stratégie de génération de tests

Moodle ou en développant une application spécifique, les outils et techniques sont disponibles pour mener à bien un tel projet. Nous avons étendu nos terrains d'expérimentation au second degré dès 2019 dans la cadre du projet B4MATIVE. Il est apparu que Moodle n'était pas du tout aussi populaire dans le secondaire que dans le supérieur. Les travaux conduisant à l'automatisation complète de la construction des tests de révisions, amorcés dans le cadre de la thèse de Denis Ollivier, se sont donc appuyés sur une stratégie consistant à développer une application dédiée restant interopérable avec les LMS du supérieur comme avec les Environnements Numériques de Travail (ENT) déployés dans le secondaire.

[QR 3.1] La génération automatique du feedback textuel dans les tests de révisions autocorrectifs Les dispositifs *Note/Rationale as feedback* et *Processus en N Phases* recyclent les contributions des étudiants réalisées pendant les séquences d'évaluations formatives orchestrées en amont par l'enseignant. Cette approche permet donc de fournir automatiquement, sans effort supplémentaire de l'enseignant, des feedbacks textuels dans les tests de révision générés.

Notons que l'expérimentation présentée, s'intéressant aux choix des différents tests de révisions en fonction de la présence ou non de feedback textuel, n'a pas considéré les tests avec feedback fourni par l'enseignant. Nous avons délibérément écarté cette option, car notre objectif était (1) d'automatiser autant que faire se peut la génération du feedback pour chaque question et (2) d'inciter les étudiants à produire davantage d'explications pendant le cours. Fournir des tests avec feedback fourni par l'enseignant implique d'une part que l'enseignant a saisi (manuellement) l'ensemble du feedback et d'autre part que les explications produites par les étudiants ne sont plus valorisées, car remplacées par la saisie de l'enseignant. Ces deux implications rentraient directement en conflit avec nos objectifs initiaux.

Les retours fournis par les enseignants à l'issue des expérimentations menées dans le second degré à partir de 2019 nous ont conduits à mitiger cette approche : au collège, la trace écrite produite ou validée explicitement par l'enseignant paraît indispensable pour compléter les feedbacks produits par les élèves.

L'évaluation formative en dehors de la classe Les tests autocorrectifs de révisions construits sur la base des travaux présentés dans cette section s'insèrent directement dans 3 des 5 stratégies caractérisant l'évaluation formative dans la théorie de BLACK et WILIAM (2009). En effet, les dispositifs *Note/Rationale as feedback* proposent des feedbacks textuels perçus comme utiles par les étudiants pour progresser dans leur apprentissage (3^e stratégie) : les étudiants ont préféré utiliser les tests de révision proposant les feedbacks textuels produits par les pairs. En fondant les feedbacks sur les productions des étudiants, ces dispositifs s'inscrivent dans la démarche d'utilisation des pairs comme

ressources d'instruction (4^e stratégie). Enfin, les tests de révisions, étant mis à disposition des étudiants en libre-service de telle sorte qu'ils puissent les mobiliser à tout instant et depuis n'importe où, contribuent à renforcer la responsabilité des apprenants sur leur propre apprentissage (5^e stratégie).

L'alignement constructiviste SCOLLER (1998) rappelle que l'évaluation influence considérablement l'apprentissage : la manière d'évaluer un étudiant façonne la quantité de choses qu'il va apprendre, les contenus qu'il va apprendre, ainsi que la manière dont il va les apprendre. La plupart des étudiants concentrent leurs efforts dans l'acquisition des connaissances et des aptitudes cognitives sur lesquelles ils seront évalués. BIGGS et al. (2022) indiquent que les pédagogies dites actives insérant l'étudiant dans des activités cohérentes avec les objectifs pédagogiques d'une part, et avec les tâches d'évaluations d'autre part, favorisent une approche d'apprentissage en profondeur quel que soit le profil de l'apprenant. Les deux auteurs mettent en avant le rôle des évaluations formatives apportant à l'étudiant un feedback tout au long de son apprentissage lui permettant de savoir où il en est par rapport aux connaissances ou compétences visées.

Les dispositifs présentés dans cette section induisent une continuité entre les concepts et notions présentés en classe, les activités d'évaluations formatives menées en classe, les tests de révisions s'appuyant directement sur les contenus produits pendant ces séquences préalables et les évaluations sommatives construites en partie sur l'ensemble de ces activités. À ce titre, les dispositifs *Note/Rationale as feedback* offrent aux enseignants et aux apprenants un soutien à l'alignement constructiviste de BIGGS.

Publications Les travaux présentés dans cette section ont été publiés dans les articles ayant conduit à la conception du processus en N Phases (SILVESTRE, 2015 ; SILVESTRE et al., 2014a, 2015a, 2015b, 2015c, 2017).

(Pas de) Rétrospective Ces travaux ayant été réalisés dans le cadre de ma thèse, ils ont déjà fait l'objet d'une rétrospective dans le chapitre 2.

3.4 La conception d'un système de révision

3.4.1 Méthode de conduite de la recherche

Les travaux s'inscrivent dans une approche de recherche fondée sur la conception (DBR pour *Design-based research*) (ANDERSON & SHATTUCK, 2012). Cette approche consiste dans le contexte de notre projet à :

- s'inscrire dans une succession de cycles conception / développement / expérimentation / analyse de résultats permettant d'améliorer la plateforme et d'en tirer des

principes de conception réutilisables et partageables.

- expérimenter la plateforme conçue en contexte écologique (avec de “vrais” élèves et de “vrais” enseignants)

Les études qualitatives et quantitatives articulées selon les plans séquentiels exploratoires et/ou explicatifs (CRESWELL & CLARK, 2017) permettront d'alimenter les différentes activités du DBR en plaçant l'humain au centre des différentes itérations sur la conception.

Ces travaux se déroulent dans le cadre de la thèse CIFRE de Denis Ollivier en partenariat avec l'entreprise Kosmos². Au moment de la rédaction de ce manuscrit, la thèse co-dirigée par Julien Broisin et moi-même, co-encadrée par Jean-Baptiste Raclet est à mi-parcours. Nous décrivons dans cette section, les premiers résultats issus de ces travaux.

3.4.2 L'étude qualitative exploratoire

Durant la première itération, une étude exploratoire a été menée auprès d'enseignants et d'étudiants visant à définir les hypothèses à mettre à l'épreuve dans le cadre d'une étude quantitative à suivre. Cette démarche correspond à ce que CRESWELL et CLARK nomment *Exploratory sequential design*.

La démarche exploratoire séquentielle se déroule en trois phases. (1) La collecte et l'analyse de données qualitatives, servent (2) à développer un outil qui sera (3) testé quantitativement. Cela signifie que l'outil sera fondé en prenant en compte les points de vue des participants. Plus largement l'approche consiste à créer un instrument (un système logiciel dans notre cas) qui sert d'étape intermédiaire entre les phases qualitatives et quantitatives. Cette approche est parfois appelée *Instrument development model* (CRESWELL et al., 2004).

En conséquence, l'étude qualitative présentée ici vise à répondre aux questions de recherches suivantes :

[QR 4.1] Quelles hypothèses permettent de guider la conception d'un outil de révision répondant aux besoins des enseignants ?

[QR 4.2] Quelles hypothèses permettent de guider la conception d'un outil de révision adapté aux besoins des élèves ?

La contribution attendue dans ce travail est bien un ensemble d'hypothèses à formuler pour guider la conception du système de révisions cible répondant aux besoins des enseignants et des apprenants. La première version du système doit permettre de collecter les données d'usage permettant de valider ou d'invalider les hypothèses via une étude quantitative.

2. <https://www.kosmos-education.com/>

3.4.2.1 Participants et collecte des données

Participants Huit enseignants bénévoles de collège et de lycée qui avaient déjà été formés à l'utilisation d'Elaastic ont été interviewés. Seuls six enregistrements valides ont pu être conservés pour cette étude. En effet, un enregistrement endommagé, ainsi qu'un entretien pour lequel nous n'avons pas reçu d'autorisation écrite d'utiliser les données, ont dû être exclus des résultats. Des entretiens individuels ont été menés auprès de 4 élèves qui venaient de terminer leurs études secondaires.

Collecte de données L'étude qualitative a consisté en des entretiens semi-structurés d'environ 45 minutes. Les questionnaires utilisés lors des entretiens ciblaient trois objectifs : la compréhension de l'environnement de travail, la présentation d'un prototype d'outil de révision et l'obtention de réponses sur l'utilisation potentielle de cet outil. Les questions permettant de répondre aux questions sur l'acceptation de l'outil ont été bâties sur les variables du modèle UTAUT2 (VENKATESH et al., 2012). Les données sur les profils des enseignants ont été collectées via une enquête en ligne préalable. Le tableau 3.1 synthétise les informations de profil sur les enseignants. Le questionnaire utilisé au cours des entretiens est présenté dans le tableau 3.2.

TABLE 3.1 – Profil des enseignants participant à l'étude

Identité			Expérience professionnelle				Aisance			
Id	Âge	Genre	Années	Disciplines	Niveaux enseignés	Nb élèves	Internet	Ordinateur	Smartphone	Tablette
En1	44	F	18	SVT	5e, 4e, 3e	315	À l'aise	Ni à l'aise ni mal à l'aise	Mal à l'aise	À l'aise
En2	48	F	23	Histoire, EMC, Géographie	6e, 5e, 4e, 3e, 2nde	170	À l'aise	À l'aise	Pas du tout à l'aise	Ni à l'aise ni mal à l'aise
En4	59	M	38	Mathématiques, Informatique	6e, 5e, 4e, 3e, 2nde, 1re, Term.	150	À l'aise	À l'aise	À l'aise	À l'aise
En5	35	M	12	Mathématiques, Informatique	6e, 5e, 4e, 3e, 2nde, 1re, Term.	150	Très à l'aise	Très à l'aise	À l'aise	À l'aise
En6	43	F	20	Mathématiques, Informatique	6e, 5e, 4e, 2nde, 1re, Terminale	150	À l'aise	À l'aise	Ni à l'aise ni mal à l'aise	Ni à l'aise ni mal à l'aise
En8	49	M	23	Physique-Chimie	6e, 5e, 4e, 3e, 2nde	300	Très à l'aise	Très à l'aise	À l'aise	À l'aise

Les informations de profils des élèves ont été collectées durant les entretiens. Les tableaux 3.3 et 3.4 présentent respectivement les données de profils des élèves et le questionnaire utilisé lors des entretiens menés auprès de ces derniers.

Le prototype non fonctionnel à haute fidélité présenté lors des entretiens a été conçu dans l'idée de proposer des contenus déjà travaillés en classe grâce à l'utilisation de la plateforme Elaastic. Le prototype a été conçu suivant une approche *mobile first*. La figure 3.6 présente les 4 écrans principaux du prototype.

Une analyse thématique a été réalisée à partir des transcriptions des entretiens.

TABLE 3.2 – Questionnaire pour les enseignants

Questions posées
Parmi vos pratiques pédagogiques, en utilisez-vous une plus fréquemment que les autres ?
Donnez-vous du travail hors classe à vos élèves ? Si oui, sous quelle forme ?
Comment demandez vous aux élèves de préparer une évaluation sommative ?
Êtes-vous à l'aise avec l'utilisation des outils numériques pour l'enseignement ? Lesquels utilisez-vous ?
Utilisez-vous des réseaux sociaux pour échanger avec les élèves ? Et pour échanger avec les autres enseignants ?
— Présentation du prototype —
Pensez-vous qu'une telle application soit pertinente pédagogiquement pour les élèves ? Pour les enseignants ?
Auriez vous des réserves par rapport à l'usage d'un tel outil ? Vis-à-vis des enseignants ? Vis-à-vis des élèves ?
Quels indicateurs pourraient vous être utiles dans une telle application ?
Voudriez vous intervenir en amont d'une telle application ? Comment ?

TABLE 3.3 – Profil des élèves participant à l'étude

Élèves	Âge	Genre	Spécialisation
El1	17	M	Physique, SVT
El2	17	F	Maths, Physique
El3	17	M	Maths, Informatique
El4	17	F	Arts Plastiques, Littérature

3.4.2.2 Résultats et hypothèses induites

Les résultats de l'étude qualitative révèlent trois thèmes majeurs sur lesquels nous avons pu formuler les hypothèses guidant la conception de la première version de l'outil.

Smartphone ou pas Smartphone ? Les enseignants et les élèves souhaitent que l'outil soit accessible sur divers appareils, pas uniquement sur les smartphones. Les enseignants sont préoccupés par une utilisation équitable et prudente des smartphones en éducation. Certains élèves ont indiqué préférer travailler sur un autre dispositif que leur smartphone afin d'éviter les distractions.

Sur la motivation L'utilisation d'outils numériques est généralement bien accueillie, car elle est perçue par les enseignants comme source de motivation pour les élèves. Certains enseignants voient la nécessité d'une *carotte* sous forme de note pour stimuler l'utilisation. De leur côté, certains élèves mentionnent le caractère décourageant de la note chiffrée quand elle est mauvaise et non accompagnée d'un feedback plus détaillé.

Du suivi des enseignants sur les travaux de révisions Enfin, les enseignants souhaitent des indicateurs pour suivre l'utilisation par les élèves, malgré des préoccupations

TABLE 3.4 – Questionnaire pour les élèves

Questions posées	
	Quel âge avez-vous ?
	A quel genre vous identifiez-vous ?
	Quelle filière et/ou spécialisation suiviez-vous au lycée ?
	Durant l'année scolaire passée, aviez-vous accès aux outils suivants le soir chez vous ?
	Par quel(s) moyen(s) échangez-vous avec vos enseignants l'an dernier ?
	Avec vos camarades de classe ?
	Quel usage faites-vous de la technologie ?
	Parmi les usages scolaires, vous en servez-vous directement en classe également ?
	Si oui, avez-vous déjà utilisé des outils en particulier dont vous vous rappelez ?
	Utilisez-vous déjà une application de révisions ? Si oui laquelle ?
	Pouvez-vous la décrire ? Si non voyez-vous à quoi cela ressemble ?
	A quoi devrait ressembler une application de révisions pour vous ?
	Quels outils vous vous attendriez à ce qu'elle vous propose ?
<i>— Présentation du prototype —</i>	
	Est-ce que vous pensez qu'une application de révisions comme celle-ci serait utile dans votre travail ?
	Si vous utilisiez une application de révisions, pensez-vous obtenir de meilleurs résultats scolaires ?
	En moyenne combien de temps par semaine utiliseriez-vous une telle application ?
	Quels éléments pourraient vous freiner dans l'utilisation de l'application présentée ?
	Quels éléments aimeriez-vous voir dans une telle application ? Pour vous aider à mieux travailler ?
	Pour apprécier l'outil / être motivé à s'en servir ?
	Voudriez-vous que vos parents puissent voir le travail fait sur cette application ?
	Pensez-vous être plus motivé à utiliser l'application si vous pouvez interagir avec les autres élèves dessus ?
	Quel rôle devraient avoir les enseignants autour de cette application d'après vous ?
	Quelles informations accepteriez-vous de partager avec l'enseignant ?
	Automatiquement ou à la demande (avec une action de votre part) ? Nominativement ou anonymement ?

sur la charge de travail supplémentaire que cela pourrait représenter. Les étudiants accueillent également favorablement un suivi personnalisé.

Ces résultats nous ont conduits à formuler les 5 hypothèses synthétisées dans le tableau 3.5.

TABLE 3.5 – Hypothèses identifiées à l'issue des entretiens

	Hypothèses
H1	Les élèves accéderont au système de révisions à l'aide d'une variété d'appareils (smartphones, ordinateurs, tablettes).
H2	Fournir aux élèves un outil de révision numérique qui exploite l'effet test aura déjà un effet bénéfique sur leur utilisation du test en tant que méthode de révisions.
H3	L'utilisation du système de révision sera particulièrement marquée à proximité des évaluations sommatives.
H4	Les enseignants utiliseront les données nominatives sur le travail des élèves pour améliorer le suivi individualisé malgré la charge de travail que cela représente.
H5	Les élèves et leurs parents accepteront que des données personnelles soient collectées afin d'élaborer des indicateurs pour les enseignants.

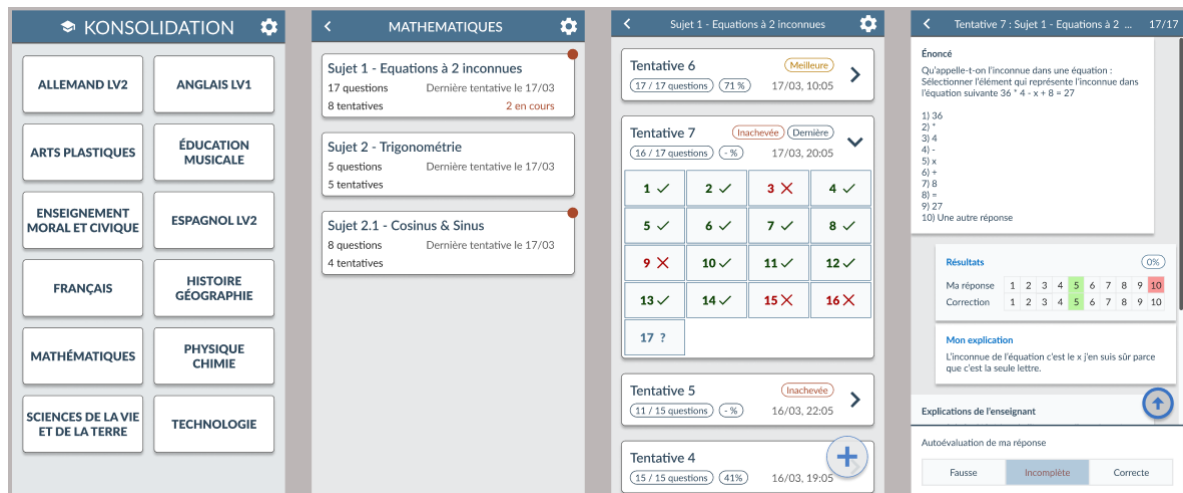


FIGURE 3.6 – Prototype haute fidélité présenté durant les entretiens : les 4 écrans principaux

3.4.2.3 L'application Konsolidation

Les hypothèses formulées ont guidé la conception de la première version de l'application de révision baptisée *Konsolidation*. Les hypothèses induisent des exigences fonctionnelles devant être satisfaites par l'application afin d'être en mesure de valider ou d'invalider ces hypothèses. À titre d'exemple, pour être capable de vérifier l'utilisation de différents types de dispositifs, l'application doit collecter les informations relatives aux dispositifs utilisés pour se connecter à l'application. Les impacts de chaque hypothèse sur la conception sont synthétisés dans le tableau 3.6.

À l'heure où ces lignes sont rédigées, la première version de l'application *Konsolidation* est en cours de développement. La mise en exploitation pour les premières expérimentations en contexte écologique est prévue pour le début d'année 2024.

3.4.3 Discussion et positionnement des travaux

Une première itération centrée sur l'effet test Les hypothèses formulées dans cette première itération nous ont conduits à implanter une première version du système cible offrant la possibilité aux élèves de se tester. Aucune incitation à espacer les séances de révisions n'a été spécifiée. En effet, notre parti pris est de faire un état des lieux objectif sur les pratiques des élèves : préfèrent-ils se tester quand on leur facilite l'accès aux tests ? Certains parviennent-ils d'eux même à planifier des séances espacées alors que notre hypothèse stipule que les élèves se contentent de réviser à l'approche de l'évaluation sommative ? Nous souhaitons en savoir davantage sur les pratiques réelles avant d'introduire des incitations à l'espacement des séances. Que l'on s'intéresse à l'activité de révisions (tests *vs* lectures répétées) ou à la planification des révisions dans le temps (espacées ou massées à l'approche d'un examen), on s'intéresse finalement à des dimensions de la

TABLE 3.6 – Impact des hypothèses identifiées lors des entretiens sur la conception de l'application Konsolidation

Hypothèses	Impacts sur la conception
H1	I1.1 L'application doit être conçue de manière à être utilisable via n'importe quel type de dispositif (<i>responsive design.</i>) I1.2 L'application doit capturer et enregistrer les données relatives aux dispositifs utilisés par les élèves (pose de sondes logicielles).
H2	I2 Aucune incitation à l'utilisation de l'outil de révision (notifications, ludification, etc.) n'est implantée dans la première version de l'application
H3	I3.1 Aucune incitation à l'espacement des séances de révision (notifications, ludification, etc.) n'est implantée dans la première version de l'application I3.2 Un dispositif de collecte de dates des évaluations sommatives doit être mis en place.
H4	I4.1 Un tableau de bord présentant des informations sur les travaux de révision des élèves doit être accessible aux enseignants. I4.2 L'application doit capturer et enregistrer les données relatives à l'utilisation des tableaux de bord par les enseignants (pose de sondes logicielles).
H5	I5 L'application permettra aux élèves et parents de sélectionner les données à caractère personnel pouvant être présentées aux enseignants dans les tableaux de bord (<i>Privacy by design</i>)

régulation de l'apprentissage.

Des tests autocorrectifs à la régulation des apprentissages Parmi les hypothèses formulées dans le cadre de ce travail, la plupart ne sont pas très originales : introduire de la flexibilité dans les dispositifs utilisables et supposer que les élèves se connecteront à la plateforme avec plusieurs types de dispositifs ne nécessitait peut-être pas une étude qualitative. Néanmoins, nous avons été surpris par la convergence des enseignants et des élèves interviewés vers une demande de régulation des apprentissages par les enseignants sur les activités de révisions (H4, H5).

Le lien entre évaluation formative et régulation des apprentissages est mis en lumière dans différents travaux (LOPEZ, 2017 ; PERRENOUD, 1998). Les tests autocorrectifs, parce qu'ils relèvent de l'évaluation formative, sont liés à la régulation des apprentissages. En permettant un travail en autonomie des apprenants, les tests autocorrectifs ont, a priori, plus à voir avec l'autorégulation qu'avec une régulation externe des apprentissages. Les résultats des entretiens tendent à faire penser que les activités de révisions au collège et au lycée gagneraient en efficacité si elles faisaient l'objet de régulations externes par les

enseignants. Un outil numérique de révisions pourrait alors être conçu comme un dispositif de transition entre apprentissage régulé et apprentissage autorégulé. Il permettrait de maintenir et de réduire progressivement la régulation externe dans les travaux hors classe ayant vocation, à terme, à se dérouler en totale autonomie.

Cette *macro-hypothèse* stipulant qu'un système de révision peut constituer un outil d'aide au développement des compétences d'autorégulation constitue à elle seule au moins un sujet de thèse. Denis abordera ce questionnement durant la fin de son doctorat s'il dispose du temps nécessaire. Une thèse à suivre sur cette problématique, toujours en partenariat avec Kosmos, fait actuellement l'objet de discussions.

Publications Les résultats de l'étude qualitative ont été publiés à la conférence EIAH 2023 (OLLIVIER et al., 2023b) en papier court et à la conférence EC-TEL 2023 (OLLIVIER et al., 2023a) en papier long. Nous avons été agréablement surpris de l'accueil de ces travaux dans les communautés nationales et internationales en raison du faible taux d'études qualitatives introductives d'une recherche plus vaste dans les conférences ciblées.

3.4.4 Rétrospective

Les travaux présentés dans cette section ne sont que la première partie de la thèse de Denis. La rétrospective est donc partielle au regard de l'objectif principal que représente la conception d'un système de révisions.

Enfin un DBR qui tienne la route! La thèse de Denis a démarré en février 2022. Rialy terminait sa thèse, et il était important de prendre en compte les points à améliorer à l'issue des travaux conduits durant cette première thèse co-dirigée. Nous voulions améliorer notre approche DBR en impliquant au plus tôt les enseignants et les élèves dans le processus de recherche. L'étude qualitative proposée a permis de co-construire avec les premiers intéressés les bases de la recherche menée par Denis.

Un renfort en ingénierie bienvenu Le partenariat avec Kosmos a permis de financer un alternant sur 2 ans inscrits en Master MIAGe ainsi que 6 mois de prestations d'ingénieur senior pour garantir la qualité des systèmes logiciels développés. Ce renfort en ingénierie avait été identifié comme nécessaire dans l'analyse de risque conduite en amont du projet : nous partions de zéro sur la conception du système cible (la thèse de Rialy s'appuyait sur le système Elaastic) et nous voulions introduire des études qualitatives dans les travaux de recherche. Ce renfort a permis effectivement de conduire des travaux de développements au plus tôt durant la thèse de Denis en parallèle de l'étude qualitative.

Une approche qualitative très gourmande en temps dans le contexte d'une thèse en informatique La conduite de 11 entretiens au total, la transcription et l'ana-

lyse des résultats ont occupé environ 6 mois de la thèse de Denis. Sur un travail de 3 ans devant déboucher sur des principes de recommandation d'un système de révisions tirant parti des effets tests et d'espacement, cette durée nous paraît trop importante. J'envisage, pour les études qualitatives à suivre deux options : (1) de privilégier, pendant le déroulement de la thèse en informatique concernée, des focus groupes en espérant que ceux-ci permettent de réduire le coût en temps des études qualitatives ; (2) de mobiliser d'autres ressources en amont de la thèse (stages de master en sciences de l'éducation ou psychologie par exemple) pour mener les études qualitatives exploratoires.

3.5 Conclusion

Nous avons présenté dans ce chapitre les travaux en lien avec les activités de révisions.

La première partie décrit les travaux menés durant ma thèse traitant la question de la génération automatisée des tests de révisions alors que la deuxième partie traite de la conception d'un système complet de révisions.

Les travaux conduits sur la génération automatisée des tests de révisions proposent une approche originale de recyclage d'une part des questions posées en face à face durant des séquences d'évaluation formative et, d'autre part des meilleures explications fournies par les étudiants durant ces activités. Ces contributions des étudiants sont alors utilisées comme source de feedbacks textuels dans les tests de révisions générés par les systèmes élaborés (Tsaap-Notes et Elaastic).

Les travaux portant sur la conception d'un système de révision se déroulent dans le cadre de la thèse de Denis Ollivier en cours au moment de la rédaction de ce document. Ils reprennent les principes de construction automatiques des tests de révisions initiés durant ma thèse en cherchant à tirer parti des effets tests et d'espacement dont les bénéfices sur l'apprentissage ont été mis en lumière par les travaux en psychologie cognitive. Seuls les résultats de l'étude qualitative menée en première partie de la thèse de Denis ont été présentés dans ce chapitre.

Du point de vue méthodologique, le chapitre présente le glissement progressif d'une approche de la recherche inspirée des méthodes de conception issue de l'industrie du logiciel (méthodes agiles et *Lean Startup*) vers une démarche itérative et incrémentale pensée par la recherche et pour la recherche : le *Design Based Research*.

L'ensemble de ces travaux ainsi que les travaux portant sur la conception des systèmes d'évaluation formative présentés dans le chapitre 2 ont fait germer une grande quantité de questions permettant de construire un programme de recherche pour un certain nombre d'années à venir.

Deuxième partie

Le projet de recherche

Chapitre 4

Retour sur la situation prototypique d'une recherche en IEAH

Contenu

4.1	De l'artefact à plusieurs instruments	74
4.2	Proposition d'une nouvelle situation prototypique	75
4.3	Elaastic passé au crible de la nouvelle situation prototypique	78
4.4	Conclusion	79

4.1 De l'artefact à plusieurs instruments

Comme évoqué dans l'introduction, TCHOUNIKINE (2009, p. 21) précise que la situation prototypique d'une recherche en IEIAH est « la conception d'un logiciel pour une situation pédagogique cible ». La figure 4.1 illustre de manière rudimentaire cette situation prototypique. La situation pédagogique cible mobilise le système informatique. Ce dernier soutient la situation cible.

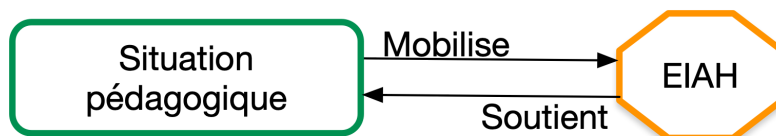


FIGURE 4.1 – Représentation de la situation prototypique proposée par TCHOUNIKINE

Les projets présentés dans la partie I s'inscrivent chacun à leur manière dans ce cadre proposé par TCHOUNIKINE. Elaastic est le résultat de la conception d'un logiciel soutenant l'évaluation formative dans le contexte de cours en face-à-face dispensés en amphithéâtre ; Konsolidation est le fruit de la conception d'un logiciel soutenant l'activité d'évaluation formative réalisée hors classe, dans le contexte de révisions, sur des notions travaillées en formatif avec une plateforme de type Elaastic.

Elaastic, dont la conception a débuté en 2012 dans le cadre de ma thèse, baptisé à cette époque Tsaap-Notes et mis entre les mains de véritables utilisateurs dès 2013, a très rapidement été utilisé par des enseignants dans des contextes n'ayant rien à voir avec la situation pédagogique initiale : l'application a été utilisée en séances de travaux pratiques de programmation pour « forcer » les étudiants à prendre une posture réflexive sur ce qu'ils codaient (RACLET & SILVESTRE, 2020); d'autres enseignants ont utilisé la plateforme pour faire travailler les étudiants, en dehors des cours, sur la compréhension de concepts en suivant le processus implanté dans Tsaap-Notes (PARMENTIER & SILVESTRE, 2019). Enfin, aujourd'hui le système est autant utilisé dans l'enseignement secondaire que dans le supérieur dans des situations pédagogiques très éloignées de la situation décrite dans les premiers travaux (ANDRIAMISEZA, 2022). C'est sans aucun doute une belle illustration du phénomène de genèse instrumentale (RABARDEL, 1995, p. 109) dans laquelle l'artefact technologique initial ne devient instrument qu'après appropriation par ses utilisateurs. Je me suis interrogé sur ce qui avait pu conduire Tsaap-Notes puis Elaastic à se décliner en autant de différentes formes d'instruments. Cette réflexion m'a conduit à proposer une nouvelle situation prototypique de la recherche en IEIAH.

4.2 Proposition d'une nouvelle situation prototypique

Je lis depuis plus d'une dizaine d'années beaucoup d'ouvrages présentant des synthèses de résultats de travaux issus des champs de la psychologie cognitive et des neurosciences. Ces lectures peuvent être positionnées suivant deux principales dimensions : la discipline dominante (allant de la psychologie aux neurosciences), et le domaine d'étude dominant (allant des émotions à la cognition). La figure 4.2 présente les auteurs et concepts associés les plus influents sur la réflexion menée dans cette section. Il a été long et difficile de tisser des liens entre ces lectures et mes travaux de recherche. Et, d'une certaine manière, ce travail d'obtention de l'HDR clôture ce cheminement.

Je me suis d'abord beaucoup intéressé à la manière dont les humains prennent des décisions (KAHNEMAN, 2016; KAHNEMAN et al., 2021; TVERSKY & KAHNEMAN, 1974), ou sur la manière dont on peut influencer et aider les humains à prendre de meilleures décisions (HOUDÉ, 2019a; THALER & SUNSTEIN, 2022). Ces lectures m'ont familiarisé avec la théorie des systèmes 1 et 2 de la pensée décrite initialement dans les travaux de KAHNEMAN. Mobilisé dans le contexte d'une prise de décision, le système 1 est rapide, économe en énergie, englobe les réactions émotionnelles, mais est aussi le père de tous les biais cognitifs. Quand il ne donne pas satisfaction, le système 2 prend le relai en utilisant le raisonnement et la logique, mais il est lent et coûteux en énergie. Le point suscitant le plus mon attention peut se résumer ainsi : la manière dont le cerveau, dans certaines situations, se met en mode « réflexion » (système 2) au détriment du mode automatique (système 1). L'ensemble de ces travaux relèvent davantage du champ de la psychologie et

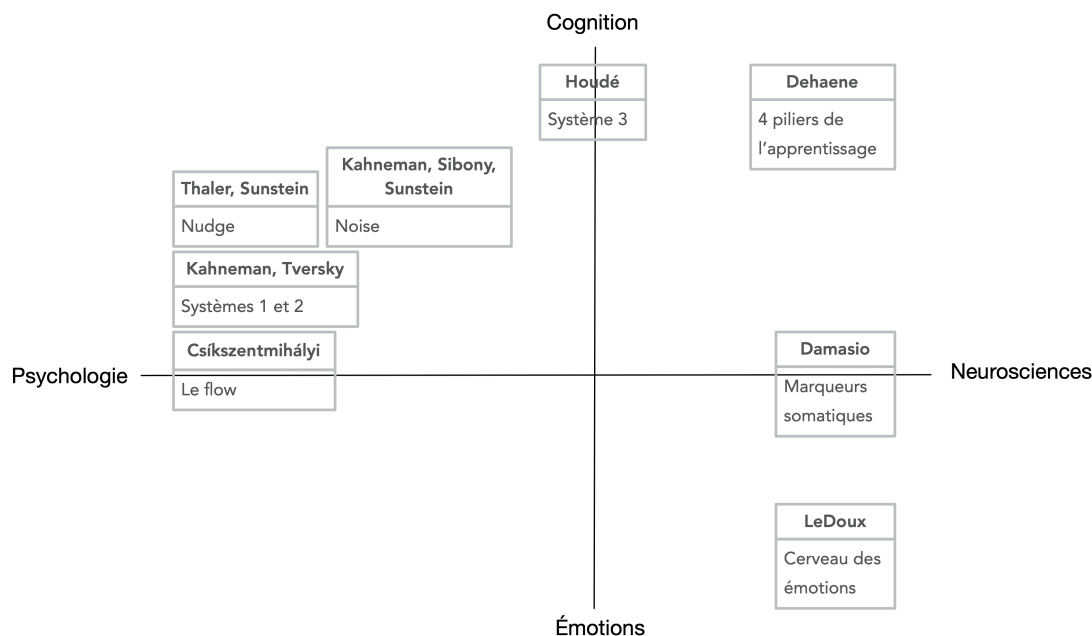


FIGURE 4.2 – Auteurs et concepts positionnés suivant 2 dimensions : la discipline dominante (de la psychologie aux neurosciences) et le domaine d'étude dominant (des émotions à la cognition)

étudient principalement des états cognitifs liés à la prise de décision.

Les travaux présentés dans (HOUDÉ, 2019a, 2019b) se situant à la frontière de la psychologie et des neurosciences montrent d'une part que le système 2 s'active quand le système 1 est inhibé par ce que HOUDÉ a désigné « système 3 », et, d'autre part que cette « résistance cognitive » est souvent nécessaire à l'acquisition des savoirs dispensés à l'école. Les travaux de HOUDÉ et plus généralement les travaux menés au Laboratoire de Psychologie du Développement et de l'Éducation de l'Enfant (LabPsyDé¹) dirigé aujourd'hui par Grégoire Borst représentèrent, pour moi, le maillon manquant liant la théorie des systèmes de pensées de KAHNEMAN avec les recherches sur l'apprentissage humain. La présentation des 4 piliers de l'apprentissage dans (DEHAENE, 2013) a permis de faire un premier lien direct avec les travaux menés sur les systèmes d'évaluation formative. Il fut assez immédiat d'établir un lien entre Elaastic et 3 des 4 piliers : Elaastic soutient des activités dans lesquelles les apprenants doivent être attentifs (1^{er} pilier), engagés cognitivement (2^{ème} pilier) et reçoivent du feedback sur leurs productions (3^{ème} pilier). La consolidation (4^{ème} pilier) est directement traitée dans la thèse en cours de Denis Ollivier.

Alors que je m'intéressais en parallèle au cerveau des émotions à travers les travaux de LEDOUX (2005), la lecture de DAMASIO (2006) apportait un éclairage éblouissant sur les liens indéfectibles entre émotions et cognition. Du côté de la psychologie, CSIKSZENTMIHALYI (2013) fournissait un nouvel objet scientifique à la frontière du cognitif et de l'émotion : le *flow*.

1. <https://www.lapsyde.com/>

Les échanges avec les praticiens de l'enseignement, c'est-à-dire les enseignants du secondaire avec qui je mène des recherches collaboratives, mais aussi les inspecteurs engagés dans ces recherches, m'ont définitivement aidé à tisser les liens entre ces lectures et mes propres travaux. Un évènement a marqué un tournant : suite à une formation sur Elaastic que j'avais dispensé aux enseignants membres du groupe de travail *Lettres et Numérique* de l'Académie d'Occitanie, je partageais mon enthousiasme avec Mme Plaisant-Soler, IA-IPR de Lettres, de l'intérêt que portaient des enseignants de Lettres du secondaire à la plateforme. Elle exprima son analyse ainsi : « Elaastic est un outil qui fait réfléchir les élèves ». Le lien avec mon travail de recherche devint alors limpide : depuis le début de ma thèse, je m'intéresse à la conception d'outils qui font réfléchir les élèves, ou, dit autrement, qui mettent les élèves dans un état cognitif « système 2 » favorable à l'apprentissage dans un contexte scolaire.

Après cette réinterprétation des travaux de recherche auxquels je participe, a émergé l'hypothèse de travail suivante : la genèse « multi-instrumentale » est possible quand le système informatique ne soutient pas seulement une situation pédagogique cible mais, quand il soutient un ou des **états cognitifs et émotionnels** favorables à l'apprentissage. C'est parce qu'Elaastic facilite la mise en œuvre de situations favorisant l'activation du système 2 et la fourniture de feedbacks qu'Elaastic est utilisable dans plusieurs situations pédagogiques différentes : potentiellement, toutes celles qui requièrent l'explicitation d'un raisonnement logique caractérisant le système 2. La situation prototypique alternative à celle proposée par TCHOUNIKINE, p. 21 pour la recherche en IEIAH est, en conséquence, la suivante : la conception d'un système informatique pour soutenir, respectivement limiter, les états cognitifs ou émotionnels favorables, respectivement défavorables, à l'apprentissage. La figure 4.3 illustre cette nouvelle situation : plusieurs situations pédagogiques mobilisent le même système informatique. Ces situations génèrent des états cognitifs et émotionnels chez les apprenants. Parmi ces états, le système informatique est conçu pour soutenir ceux favorables à l'apprentissage et pour limiter ceux défavorables à l'apprentissage.

Cette situation offre un cadre de travail pour les recherches en IEIAH s'appuyant sur les résultats de la recherche en psychologie cognitive et en neurosciences de l'éducation. Elle doit être accompagnée, pour ne pas être complètement hors sol par rapport aux problématiques de terrain, de situations pédagogiques servant (1) à illustrer l'utilisation du système conçu, (2) à évaluer en contexte écologique le système. Les travaux de CHI et al. (2018) sont à ce titre très inspirants : il s'agit bien de transférer dans les pratiques, à l'aide d'un logiciel dans notre cas, le soutien d'un ou plusieurs états cognitifs ou émotionnels favorables à l'apprentissage.

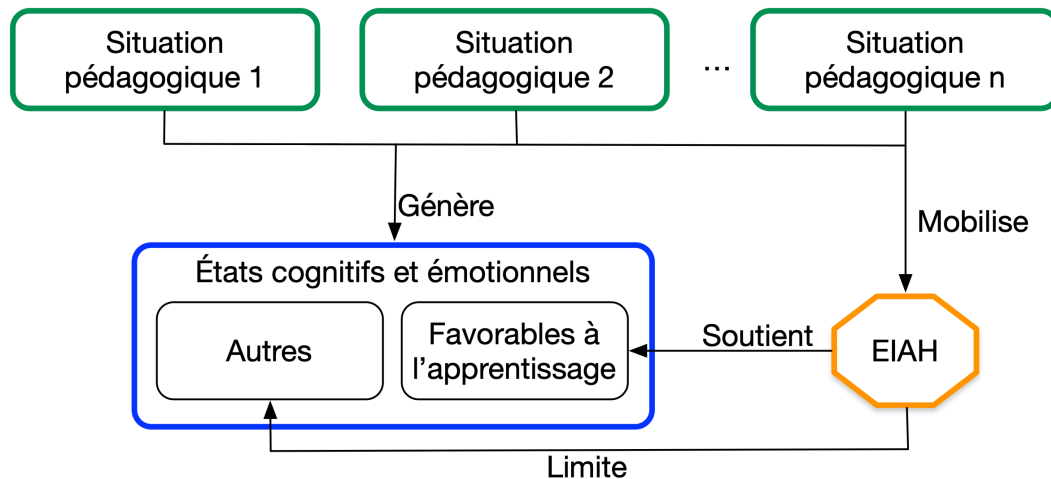


FIGURE 4.3 – Représentation de la nouvelle situation prototypique

4.3 Elaastic passé au crible de la nouvelle situation prototypique

Nous proposons, dans cette section, de spécifier Elaastic en termes de soutien et de limitation à des ensembles d'états cognitifs et émotionnels des apprenants respectivement favorables et défavorables à l'apprentissage. Le tableau 4.3 présente une synthèse des liens entre les interactions proposées par Elaastic et les impacts visés sur les états cognitifs ou émotionnels.

TABLE 4.1 – Liens entre interactions Elaastic et états cognitifs/émotionnels impactés.

Interaction proposée par Elaastic	États cognitifs et/ou émotionnels cibles	Impact	Cadre
Élaboration d'une explication	Système 2	Soutien	Systèmes 1 et 2
	Attention, engagement	Soutien	Piliers de l'apprentissage
Évaluation par les pairs	Système 1	Limitation	Systèmes 1 et 2
	Génération de feedbacks	Soutien	Piliers de l'apprentissage
Restitution des résultats d'évaluations	Réception de feedbacks	Soutien	Piliers de l'apprentissage

Elaastic, en demandant explicitement aux étudiants d'élaborer des explications dans la réponse qu'ils fournissent à une question, soutient les 2 premiers piliers de l'apprentissage de DEHAENE (attention et engagement actifs) et soutient la mobilisation du système 2 de pensée au sens de KAHNEMAN. La phase d'évaluation par les pairs, telle qu'elle est conçue (anonymat préservé, équilibre des réponses à évaluer entre correctes et incorrectes) vise à réduire les biais cognitifs d'influence relationnelle tels que le biais de conformité, de confirmation, l'effet de halo, etc. Elaastic vise ainsi à limiter le champ d'action du système 1. L'évaluation par les pairs, accompagné de l'évaluation automatique dans le cadre des questions à choix, permet à Elaastic de soutenir la production de feedbacks (3^{ème} pilier de l'apprentissage). La restitution des résultats dans le tableau de bord de séquence renforce ce soutien en facilitant la fourniture d'un feedback par l'enseignant en fin de séance.

4.4 Conclusion

Nous avons proposé dans ce chapitre une nouvelle situation prototypique d'une recherche en IEIAH : la conception d'un système pour soutenir, respectivement limiter, les états cognitifs ou émotionnels favorables, respectivement défavorables, à l'apprentissage. Une réinterprétation des travaux menés autour de la plateforme Elaastic a permis d'illustrer concrètement les potentiels apports d'un EIAH dans cette démarche de soutien ou de limitation d'états cognitifs ou émotionnels. Alors que les travaux conduits jusqu'à présent n'ont pas été explicitement fondés sur cette nouvelle vision d'une situation prototypique d'une recherche en IEIAH, le programme de recherche présenté dans le chapitre suivant s'inscrit directement dans celle-ci.

Chapitre 5

Programme de recherche

Contenu

5.1	Ingénierie des EIAH	80
5.1.1	Soutiens à l’activation du système 2 et à l’inhibition du système 1	80
5.1.2	Soutien à la production de feedbacks personnalisés	82
5.1.3	Intégration des nouveaux axes de recherche	83
5.2	Méthode de conduite de la recherche	84
5.2.1	Soutien à la démarche expérimentale en contexte écologique . .	84
5.2.2	Formation et accompagnement des doctorants	85
5.3	Conclusion	87

5.1 Ingénierie des EIAH

5.1.1 Soutiens à l’activation du système 2 et à l’inhibition du système 1

Les travaux conduits autour de la plateforme Elastic notamment durant la thèse de ANDRIAMISEZA, s’intéressent exclusivement aux processus à 2 votes. Les premiers résultats de l’analyse de données exploratoire sur la répartition des séquences en fonction de son caractère bénéfique, non bénéfique ou nuisible montrent que près de 30% des séquences menées à leur terme ne conduisent pas à une amélioration du taux de bonnes réponses au second vote. Les indicateurs mis en lumière permettent donc de recommander l’abandon de la phase de confrontation de points de vue et le second vote si les chances d’obtenir une séquence bénéfique sont trop faibles. Ils permettent de flécher les réponses à commenter pendant la phase de discussions en prenant en compte les réponses erronées populaires auprès des apprenants. Ces approches restent néanmoins « prisonnières » du cadre posé par le *Peer Instruction* de MAZUR. On cherche un « traitement efficace » à proposer entre les 2 votes pour améliorer le taux de réponses, on abandonne si les conditions ne sont pas

favorables. Cette démarche nous paraît tout à fait louable dans un contexte peu informé : dans le cas du *Peer Instruction*, ni l'enseignant, ni le système n'ont connaissance nominativement de qui a répondu quoi, avec quel degré de confiance. Les systèmes implantant les processus à 2 votes dans Tsaap-Notes et Elaastic disposent de ces informations. Il nous paraît, dans ce contexte informé, pertinent d'envisager d'autres alternatives à l'abandon d'un traitement après le premier vote lorsque les conditions ne sont pas favorables.

Prenons le cas d'école exposé par les travaux de ANDRIAMISEZA et al. (2021b) justifiant l'abandon de la phase de confrontation de points de vue dans un processus à 2 votes : les étudiants ayant fourni une réponse erronée sont plus confiants que les étudiants ayant fourni la bonne réponse. Nous pouvons formuler plusieurs hypothèses à l'égard des étudiants ayant fourni la mauvaise réponse :

1. Ils sont novices sur la notion étudiée (ils se trompent).
2. Ils ont mobilisé des heuristiques de résolution de la question non adaptée et sont néanmoins satisfaits de leur réponse : le système 1 a été mobilisé.

Nous pouvons alors nous intéresser à la manière dont le système informatique, dans cette situation, peut venir remettre en question la mobilisation du système 1 et soutenir l'activation du système 2, au lieu de suggérer l'abandon de la phase de confrontation de points de vue. Par exemple, pourquoi ne pas proposer directement le choix correct aux apprenants qui se sont trompés en leur demandant de fournir une justification à ce choix ? Cette approche forcerait ces apprenants encore novices sur la notion à réfléchir sur l'exercice résolu (ATKINSON et al., 2000 ; RENKL, 2014 ; SWELLER, 2006). Nous émettons ainsi une nouvelle hypothèse : l'approche consistant à faire étudier des exercices résolus tend mécaniquement à inhiber le système 1, agissant comme un système 3 (HOUDÉ, 2019b) externalisé dans la forme de l'activité proposée. C'est donc tout un champ d'hypothèses qui s'offrent à nous pour améliorer ces processus, qui selon le contexte, resteraient bien ancrés dans une approche à 2 votes, mais qui pourraient, dans des contextes spécifiques, prendre la forme de processus plus adaptatifs garantissant de meilleurs résultats d'apprentissage en soutenant l'activation du système 2 au détriment du système 1.

Nous proposons de formuler nos premiers objectifs de recherche ainsi :

OR1.1 Améliorer le soutien à l'activation du système 2 dans les EIAH.

OR1.2 Favoriser l'inhibition du système 1 dans les EIAH.

Plus précisément, les questions de recherche que nous souhaitons explorer à court terme sont les suivantes :

QR1.1 Au sein d'un système informatique interactif, quelles sont les tâches et interactions à proposer à un apprenant en situation d'échec sur une question, un exercice ou un problème afin de soutenir le système 2 au détriment du système 1 en vue d'un meilleur apprentissage pour lui-même ?

QR1.2 Au sein d'un système informatique interactif, quelles sont les tâches et interactions à proposer à un apprenant en situation de succès sur une question, un exercice ou un problème afin de maintenir l'activation du système 2 en vue d'un meilleur apprentissage pour lui-même et les autres ?

Ces questions pourront être abordées dans la continuité de nos travaux sur les systèmes d'évaluation formative en prenant appui, pour les efforts de conception et de développement, sur la plateforme Elaastic.

5.1.2 Soutien à la production de feedbacks personnalisés

Un des moteurs de nos travaux de recherche passés et présents est le suivant : dans un contexte d'enseignement de masse nécessitant des solutions soutenables à grande échelle, la production personnalisée et fréquente de feedbacks à destination des apprenants ne peut pas être réalisée par l'enseignant seul. D'autre part, K. TOPPING (2018, p. 1) indique que les bénéfices de l'évaluation par les pairs sur l'apprentissage sont améliorés quand les évaluateurs fournissent un feedback *élaboré* composé d'une argumentation et d'exemples pouvant susciter la discussion autour de points de vue différents.

Nous pouvons ainsi formuler un nouvel objectif de recherche ainsi :

OR3 Améliorer la production de feedbacks élaborés (au sens de K. TOPPING) dans les EIAH.

Plus précisément nous souhaitons étudier la production de ces feedbacks quand elle est prise en charge par les apprenants eux-mêmes (évaluation par les pairs) ou par des Intelligences Artificielles. En conséquence, les questions de recherche que nous souhaitons aborder sont les suivantes :

QR3.1 Comment concevoir un système informatique interactif soutenant la production de feedbacks élaborés dans un contexte d'évaluation par les pairs ?

QR3.2 Comment concevoir un système informatique interactif soutenant la production de feedbacks élaborés en mobilisant les intelligences artificielles génératives et conversationnelles ?

Des travaux ont déjà été initiés en 2023 pour traiter ces 2 questions de recherche.

Amélioration de la production de feedbacks dans le contexte d'une évaluation par les pairs Le dispositif d'évaluation par les pairs proposé dans les processus d'évaluation formative soutenus par Tsaap-Notes et Elaastic est très rudimentaire : un seul critère (le degré d'accord) est proposé pour effectuer l'évaluation avec une échelle de Likert à 5 niveaux. Nous avons à cet effet initié les travaux DRAXO (SILVESTRE, 2022). DRAXO est un cadre facilitant la mise en œuvre de l'évaluation par les pairs de réponses

textuelles et qui visent la constitution par les évaluateurs de feedbacks élaborés au sens de K. TOPPING. Un focus groupe mené auprès d’enseignants et d’étudiants du secondaire et du supérieur a permis de finaliser les spécifications de DRAXO pour une implantation dans la plateforme Elaastic. Les développements ont été finalisés en octobre 2023 pour de premières expérimentations en contexte écologique prévues en novembre et décembre 2023.

L’évaluation par les pairs *vs* la correction automatique Quand j’ai débuté ma thèse en 2012, les intelligences artificielles (IA) génératives conversationnelles telles que ChatGPT¹ n’existaient pas au niveau de performance constaté aujourd’hui. Il nous paraît important d’étudier comment mobiliser cette famille d’IA dans l’enseignement de telle sorte que leurs impacts sur l’apprentissage soient positifs. Nous avons mené quelques expérimentations exploratoires sur ChatGPT 3.5 et 4 afin d’évaluer la pertinence de mobiliser l’IA dans la génération de feedbacks élaborés adressés aux étudiants. Une de ces explorations a été restituée à la Journée Pédagogique de l’Assemblée Régionale des IUT en 2023 (SILVESTRE, 2023). Au printemps 2023, nous avons initié les travaux permettant d’intégrer ChatGPT dans Elaastic afin de conduire des travaux rigoureux sur les apports de l’IA dans la fourniture de feedbacks dans le contexte d’évaluations formatives. Nous souhaitons, entre autres, comparer l’impact du feedback produit par l’IA par rapport à celui provenant des pairs. Ces travaux nous ont déjà conduits à de nouvelles questions d’ordre éthique sur la mobilisation d’IA auprès des apprenants et des enseignants : à quelles conditions l’intervention d’IA générative est-elle acceptable dans un EIAH ? Quels garde-fous mettre en place pour améliorer l’acceptabilité de systèmes incluant les IA génératives ?

5.1.3 Intégration des nouveaux axes de recherche

La figure 5.1 illustre de manière synthétique l’intégration des nouveaux axes et nouvelles problématiques de recherche en lien avec l’ingénierie des EIAH. Les lignes en pointillés sur la figure matérialisent les liens entre les nouvelles problématiques et les travaux précédents.

1. <https://chat.openai.com/>

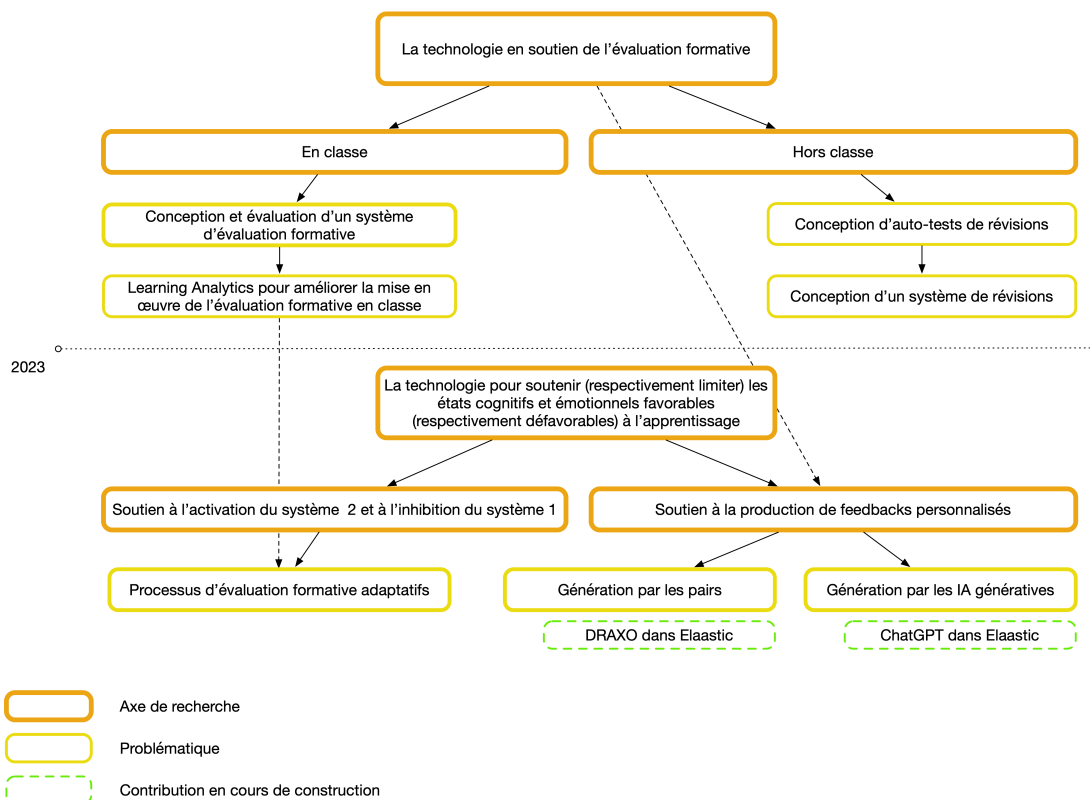


FIGURE 5.1 – Nouveaux axes et nouvelles problématiques de recherche en IEIAH.

5.2 Méthode de conduite de la recherche

5.2.1 Soutien à la démarche expérimentale en contexte écologique

En introduction de ce manuscrit (section 1.3.1), j'ai exposé mon penchant pour les approches expérimentales en revendiquant une posture post-positiviste non dogmatique. La difficulté de mettre en œuvre une approche expérimentale en contexte écologique dans une recherche en sciences de l'éducation est un verrou scientifique se situant au niveau de la méthode. Tout verrou scientifique est un candidat naturel à un projet de recherche. Nous pouvons ainsi formuler un objectif de recherche, dont les contributions porteront d'abord sur la méthode de conduite de la recherche dans le champ de l'apprentissage humain soutenu par la technologie :

OR4 Soutenir les approches expérimentales en contexte écologique dans les recherches en sciences de l'éducation en mobilisant les systèmes informatiques.

DUFLO et BANERJEE sont lauréats du prix Nobel d'économie 2019. Les travaux récompensés le furent en partie sur l'effort entrepris par DUFLO et BANERJEE pour mesurer l'impact des politiques publiques sur la pauvreté à l'aide d'approches expérimentales (DUFLO & BANERJEE, 2019). La lecture de leur ouvrage *Repenser la pauvreté*, m'a

conduit à penser qu'il devait être possible de mesurer, dans certaines situations, l'impact d'une stratégie pédagogique dans les salles de classe de la même manière.

En informatique, le principe de l'*A/B testing* est décrit comme une forme de test d'hypothèse dans laquelle deux variantes d'un logiciel sont évaluées en contexte écologique (allant de variantes avec une interface graphique légèrement modifiée à des variantes de logiciels avec de nouvelles fonctionnalités) (QUIN et al., 2023).

Nous considérons que les processus à deux votes offrent un cadre pertinent pour la mise en œuvre d'expérimentation de type *A/B testing* afin d'évaluer l'impact d'une stratégie pédagogique. Imaginons la situation suivante : nous sommes dans le contexte d'un enseignement dispensé en face-à-face dans un des amphithéâtres de médecine ou de droit contenant plusieurs centaines d'étudiants. Nous voulons comparer l'impact d'une évaluation par les pairs menée à l'aide de DRAXO avec l'impact d'une évaluation par les pairs telle qu'elle est implantée actuellement dans Elaastic. Tous les étudiants suivent le cours dispensé par l'enseignant. Au bout de quelques dizaines de minutes, l'enseignant pose une question à choix exclusif visant à évaluer la compréhension des nouvelles notions exposées. Les étudiants disposent d'une version améliorée de la plateforme Elaastic pour participer à l'activité d'évaluation formative. À l'issue du premier vote, pour la mise en œuvre de la phase d'évaluation par les pairs, la plateforme répartit les étudiants de manière aléatoire dans 2 groupes de même taille, de telle sorte que le premier groupe utilise le dispositif actuel et que le deuxième groupe utilise DRAXO. À l'issue de cette phase, les étudiants répondent une seconde fois à la même question. La plateforme peut alors, à la volée, appliquer les tests statistiques pour comparer les résultats des deux groupes d'étudiants. Dans cette expérience, les étudiants des deux groupes subissent un « traitement pédagogique » différent entre les deux votes, toutes choses étant à peu près égales par ailleurs (même enseignement dispensé par le même enseignant au même moment).

Il manque peu d'exigences à l'architecture logicielle d'Elaastic pour rendre l'expérience de pensée décrite ci-dessus réelle : il est déjà actuellement possible de choisir entre différentes modalités d'évaluation par les pairs pour une même séquence. Une fois les modalités spécifiées, elles s'appliquent pour tous les étudiants participant à la séquence. Il manque donc à Elaastic la possibilité de répartir de manière aléatoire différentes modalités de l'évaluation au sein d'une même séquence.

5.2.2 Formation et accompagnement des doctorants

LATOURE (2001) a exploré en profondeur la science « en train de se faire » dans ses travaux d'anthropologie des sciences. Dans la partie « Discussion » de son ouvrage *Le métier de chercheur : regard d'un anthropologue*, l'auteur souligne que bien des aspects de la recherche ne sont jamais enseignés aux doctorants :

« La longue épreuve initiatique imposée aux jeunes scientifiques est absurde

et cruelle. Elle consiste à passer de la maîtrise au doctorat, du doctorat au post-doc, et à découvrir par soi-même la réalité de la science qui se fait... Et beaucoup de jeunes scientifiques vivent cela très mal, je l'ai dit, comme une déchéance. "La science n'est pas ce que vous aviez dit, c'est horrible, c'est de la cuisine, il faut publier, tenir compte de l'ordre des citations, avoir des patrons qui signent parfois à votre place, il faut avoir de l'argent". »

Je n'ai personnellement pas vécu mon voyage initiatique au sein de la thèse et des quelques années suivantes comme une déchéance, mais je partage avec LATOUR l'idée que des choses qui pourraient être enseignées ne le sont pas. Ce constat mène au dernier objectif en lien avec la manière de diriger la recherche :

OR5 Améliorer la formation et l'accompagnement des doctorants.

J'ai déjà mené différentes actions pour tenter d'améliorer la situation dans le champ de la recherche sur les EIAH.

Engagements au niveau de la communauté scientifique. C'est dans cet esprit que j'ai rejoint en 2020 le groupe de travail « Épistémologie » de l'association ATIEF² regroupant la communauté francophone des chercheuses et chercheurs s'intéressant aux EIAH. Ce groupe animé par Nadine Mandran et Eric Sanchez a mené des actions de formation en direction des jeunes chercheuses et chercheurs dont la plus emblématique est sans nul doute l'école d'automne 2022 portant sur les « Méthodes de conduite de la recherche en EIAH »³.

Engagements au niveau des doctorants. Autre conséquence de cette convergence de vue avec LATOUR, j'apporte un soin particulier au suivi et à l'encadrement des doctorants sous ma responsabilité en m'inspirant des 3 formes de soutien décrites dans l'étude de DE CLERCQ et al. (2021) :

- Fournir de la structure. Cela consiste à fournir de manière régulière des informations claires aux doctorants sur les attentes du processus doctoral et la manière de le mener à bien.
- Être engagé. Cela consiste à œuvrer à une relation de qualité avec le doctorant en faisant preuve, autant que faire se peut, de chaleur et d'empathie à son égard.
- Soutenir l'autonomie. Cela consiste à accorder le degré de liberté nécessaire au doctorant afin qu'il s'approprie et co-détermine la trajectoire donnée à la thèse de doctorat.

2. <http://www.atief.fr/>

3. <https://methodo-eiah.unige.ch/>

Dans les faits, ces soutiens prennent la forme de réunions fréquentes formelles (pour la dimension « Structure ») et informelles (incarnant souvent la dimension « Engagement »). Le soutien à l'autonomie prend la forme d'une discipline de l'écoute et de l'acceptation inconditionnelle de toutes les idées provenant du doctorant qui ne rentre pas en contradiction avec les objectifs de recherche de haut niveau.

Études d'exemples résolus appliquées à la formation à la recherche. Mes travaux en lien avec les sciences de l'éducation m'ont conduit à m'intéresser à une approche pédagogique consistant à étudier des exemples résolus (*worked examples*). Un exemple résolu est composé de l'énoncé d'un problème ou d'un exercice accompagné de sa solution. L'étude d'un exemple résolu consiste, pour l'apprenant, à étudier en profondeur la solution proposée. Ce travail en profondeur est traditionnellement guidé par des questions portant sur les différentes étapes de la solution. Les résultats de la recherche démontrent la plus grande efficacité de cette approche sur les apprenants novices par rapport à l'approche consistant à faire directement résoudre des problèmes du domaine aux apprenants (ATKINSON et al., 2000 ; RENKL, 2014).

L'activité de revue de la littérature conduit à la collecte et lecture de nombreux articles sur un domaine d'intérêt donné. Il arrive alors parfois de lire un article emblématique de la recherche : un exemple de ce que l'on cherche à produire, en tant que chercheur, résolu de manière exemplaire.

J'ai souhaité exploiter ces perles scientifiques, d'abord avec les doctorants que j'encadre, en leur demandant d'étudier ces articles de manière approfondie de telle sorte qu'ils puissent s'en inspirer pour leur propre publication. Cette approche a notamment joué un rôle déterminant dans la forme de l'article « Recommendations for Orchestration of Formative Assessment Sequences : A Data-Driven Approach » ([andriamiseza_2021](#)) récompensé du prix du meilleur papier à la conférence EC-TEL 2021.

J'ai ensuite proposé et coanimé, avec Nadine Mandran, l'atelier « La diversité de la recherche en EIAH par l'exemple »⁴ lors de la conférence EIAH 2023 visant à initier la collecte de tels articles emblématiques. Plusieurs chercheurs et chercheuses ont ainsi été invités à présenter l'étude d'un article emblématique reflétant au mieux les travaux dans leur champ de recherche. Il leur était interdit de choisir une de leurs publications. Le travail de collecte et le partage de ces articles avec la communauté EIAH se poursuit.

5.3 Conclusion

Le programme de recherche présenté dans ce chapitre se développe suivant deux axes.

Le premier axe, l'ingénierie d'EIAH, est dans la droite lignée des travaux entrepris depuis ma thèse. Les objectifs sont maintenant exprimés du point de la nouvelle situation

4. <https://eiah2023.sciencesconf.org/resource/page/id/25>

prototypique présenté dans le chapitre précédent (section 4.2) : les objectifs de recherche sont formulés sous la forme de soutiens ou de limitations d'états cognitifs ou émotionnels en vue d'améliorer les apprentissages.

Le deuxième axe vise à apporter des contributions sur la méthode de conduite de la recherche dans notre domaine et plus largement dans le domaine des sciences de l'éducation. Nous voulons contribuer à l'élaboration d'instruments facilitant le développement d'approches expérimentales et à l'amélioration de la formation et de l'accompagnement des doctorants durant leur thèse.

Nous sommes partenaires du projet SCICONUM porté par la société Didask⁵ et lauréat 2023 de l'appel à manifestation d'intérêt « Compétences et métiers d'avenir » lancé dans le cadre du grand plan d'investissement « France 2030 ». Ce projet va nous permettre de financer à hauteur de 280k€ les premiers développements et les premières expérimentations à mener sur ces nouveaux axes de recherche via le recrutement de postdoctorants et d'un ingénieur de recherche. Les réponses aux projets régionaux et nationaux dans un premier temps, puis européens dans un second temps seront les leviers pour conduire nos recherches à moyen et long termes. Les partenariats déjà scellés avec les académies et les entreprises partenaires seront maintenues et renforcées afin de faciliter le transfert des résultats de recherche dans les pratiques des enseignants et des apprenants.

Les directions proposées ne représentent pas une liste exhaustive. Les échanges avec les doctorants, avec les membres de notre communauté scientifique pluridisciplinaire ainsi que toutes les lectures plus ou moins proches de notre champ de recherche sont autant de sources d'inspiration pour des questions de recherche plus intéressantes les unes que les autres. Je m'efforcerai, en dirigeant les recherches du programme proposé, d'apporter quelques éléments de réponses à certaines de ces questions.

5. <https://www.didask.com/>

Troisième partie
Rapport d'activités

Chapitre 6

Curriculum Vitae résumé

Contenu

6.1	Données personnelles	90
6.2	Parcours professionnel	90
6.3	Cursus universitaire	91

6.1 Données personnelles

Nom : **Silvestre**

Prénom : **Franck**

Date de naissance : **17/12/1972**

Situation de famille : **Marié, 1 enfant**

Grade : **Maitre de conférences, classe Normale**

Établissement : **Université Toulouse Capitole, IUT de Rodez**

Section CNU : **27**

Unité de recherche : **IRIT, UMR 5505 CNRS**

6.2 Parcours professionnel

J'ai commencé ma carrière professionnelle dans l'Éducation Nationale comme professeur de Mathématiques. Reprenant un cursus d'Informatique me menant au titre d'Ingénieur, en juin 2000 je crée la société Fylab spécialisée dans la mise en œuvre d'architectures applicatives distribuées. Fylab, après avoir fortement contribué au développement de l'Espace Numérique de Travail (ENT) de différentes Universités, développe la solution ENT Éliot pour le secondaire. La solution Éliot est retenue en 2009 par la Région Île-de-France pour équiper l'ensemble de ses Lycées. Fylab est rachetée en avril 2009 par OMT, un des

éditeurs leaders dans les solutions de vie scolaire du secondaire sur le marché français. J'ai occupé jusqu'au 31/12/2010 le poste de Directeur Technique de Fylab et d'OMT. À partir de janvier 2011, je partage mon temps de travail entre la direction de la société Ticetime co-fondée avec John Tranier et comme professeur associé à l'Université Toulouse 3 Paul Sabatier jusqu'en août 2018. C'est durant cette période que je m'inscris en thèse et initie mes travaux de recherche au sein de l'IRIT dans le domaine des Environnements Informatiques pour l'Apprentissage Humain (EIAH). Je soutiens ma thèse en novembre 2015. Je suis depuis septembre 2018 Maître de conférences en informatique : j'enseigne à l'IUT de Rodez et poursuis mes recherches au sein de l'IRIT, dans l'équipe TALENT¹.

Depuis 09/2018	Maître de conférences en Informatique Université Toulouse 1 Capitole, IUT de Rodez Laboratoire IRIT, équipe TALENT
09/2012 - 08/2018	Enseignant-chercheur associé Université Toulouse 3 Paul Sabatier Laboratoire IRIT, équipe SIERA
2011 - 2018	Directeur fondateur Ticetime, Toulouse
2009 - 2010	Directeur technique OMT, Lyon Suite du rachat de Fylab par OMT en 2009. OMT est le 2ème éditeur en France sur le marché de la vie scolaire.
2000 - 2009	Directeur fondateur Fylab, Toulouse Fylab était une ESN spécialisée dans le numérique éducatif
1994-1999	Professeur certifié de mathématiques Éducation nationale Deux ans en Côte d'Ivoire (1995-1997), puis en Alsace.

6.3 Coursus universitaire

Les mentions en gras correspondent à l'acquisition d'un titre ou d'un diplôme.

1. <https://www.irit.fr/TALENT/site/>

- 2015 Docteur en Informatique**
Université Toulouse 3 Paul Sabatier
- 2005 Ingénieur Informatique**
CNAM Toulouse
- 1999 DEA Informatique de l'Image et du Langage**
Major de promotion sur les épreuves théoriques
Université Toulouse 3 Paul Sabatier
- 1998 Maîtrise Informatique Miage**
Major de promotion
Université de Haute Alsace
- 1995 CAPES de Mathématiques**
Licence de Philosophie à l'Université Toulouse II le Mirail
- 1994 Épreuves théoriques CAPES Mathématiques**
DEUG Philosophie à l'Université Toulouse II le Mirail
- 1993 Licence de Mathématiques**
Université Toulouse 3 Paul Sabatier

Chapitre 7

Activités de recherche

Contenu

7.1	Présentation synthétique des thématiques de recherche . . .	93
7.2	Encadrements scientifiques	94
7.3	Prix et distinctions	96
7.4	Diffusion du savoir, responsabilités et activités au sein des sociétés savantes ou associations	97
7.5	Organisation colloques, conférences, journées d'étude	99
7.6	Responsabilités scientifiques	99
7.7	Liste des publications	100

7.1 Présentation synthétique des thématiques de recherche

Mes recherches s'inscrivent dans le champ des Environnements Informatiques pour l'Apprentissage Humain (EIAH), ce qui au niveau international correspond au champ Technology Enhanced Learning (TEL). Je m'intéresse plus particulièrement à la conception de systèmes informatiques qui, d'une part, engagent cognitivement les étudiants dans des activités réalisées sur ordinateur, tablette ou smartphone, et qui, d'autre part, aident les enseignants à prendre des décisions durant ou en dehors des cours. Mes travaux se déclinent sur 2 axes principaux :

- La conception de systèmes d'évaluation formative : ce fut l'objet de ma thèse soutenue en 2015 ; deux thèses en cours poursuivent les travaux dans cet axe.
- La conception de systèmes de suivi et d'évaluation de la performance d'étudiants engagés sur des tâches authentiques : ces travaux ont débuté en 2018 dans le contexte de l'enseignement de l'ingénierie du logiciel ; une thèse vient de débiter sur cet axe.

Sur le plan méthodologique, je mets en œuvre une approche de recherche fondée sur la conception (DBR pour *Design-based research*). Cette approche consiste dans le contexte des projets que je porte à :

- S’inscrire dans une succession de cycles conception / développement / expérimentation / analyse de résultats permettant d’améliorer les systèmes et d’en tirer des principes de conception réutilisables et partageables.
- Expérimenter les systèmes conçus en contexte écologique (avec de “vrais” élèves et de “vrais” enseignants)

Les études qualitatives et quantitatives articulées selon les plans séquentiels exploratoires et/ou explicatifs permettent d’alimenter les différentes étapes des cycles DBR en plaçant l’humain au centre des différents cycles de conception. L’analyse des données d’apprentissages, ou *Learning Analytics*, ainsi que l’appui sur des théories issues des champs de la psychologie cognitive ou des sciences de l’éducation prennent une place importante dans nos travaux.

7.2 Encadrements scientifiques

Thèses

Ci-dessous sont présentés les trois encadrements de thèses auxquels je suis associé.

Rialy Andriamiseza

Titre de la thèse : Learning Analytics-Based Formative Assessment Recommendations for Technology-Enhanced Learning Practices

Date de début : 1er Septembre 2019

Date de soutenance : 25 novembre 2022

Taux d’encadrement : 50%

Co-encadrant : Julien Broisin (IRIT, équipe Talent)

Financement : Projet B4MATIVE! Financé par la Direction du Numérique pour l’Éducation du Ministère de l’Éducation Nationale de la Jeunesse et des Sports.

Denis Ollivier

Titre de la thèse : Conception d'un Système Interactif pour la mise en œuvre des effets testing et spacing dans un contexte de travail scolaire hors classe.

Date de début : 1er février 2022

Date de soutenance : -

Taux d'encadrement : 50%

Co-encadrant : Julien Broisin (IRIT, équipe Talent) – 20%
Jean-Baptiste Raclet (IRIT, équipe ACADIE) – 30%

Financement : Thèse CIFRE en partenariat avec la société Kosmos¹.

Mika Pons

Titre de la thèse : Identification de patterns comportementaux des apprenants à partir de traces Git dans les activités de génie logiciel

Date de début : 1er Septembre 2021

Date de soutenance : -

Taux d'encadrement : 40%

Co-encadrant : Jean-Michel Bruel (IRIT, équipe SMART) – 20%
Jean-Baptiste Raclet (IRIT, équipe ACADIE) – 40%

Financement : Contrat ministériel.

Master 2

Ci-dessous sont présentés les trois encadrements de Master 2 auxquels je suis associé.

Mika Pons

Formation : Master 2 Informatique parcours Développement Logiciel

Titre du mémoire : Conception d'un tableau de bord pour enseignants à partir des traces d'apprenants collectées avec Git.

Période : Octobre 2020 à Juillet 2021

Taux d'encadrement : 50%

Co-encadrant : Jean-Baptiste Raclet (IRIT, équipe ACADIE) – 50%

Financement : Alternance financée par le Master Développement Logiciel de l'Université Toulouse 3 Paul Sabatier.

Rémi Saurel

Formation : Master 1 et 2 mention MIAGe
 Titre du mémoire : Conception d'une application de révisions pour les élèves du secondaire.
 Période : Septembre 2022 à Juillet 2024
 Taux d'encadrement : 50%
 Co-encadrant : Jean-Baptiste Raclet (IRIT, équipe ACADIE) – 50%
 Financement : Alternance financée dans le cadre de la thèse CIFRE de Denis Ollivier en partenariat avec la société Kosmos.

7.3 Prix et distinctions**Best Paper Award EC-TEL 2021**

ANDRIAMISEZA, R., SILVESTRE, F., PARMENTIER, J.-F., & BROISIN, J. (2021b). Recommendations for Orchestration of Formative Assessment Sequences : A Data-Driven Approach. In *Technology-Enhanced Learning for a Free, Safe, and Sustainable World* (p. 245-259). Springer International Publishing. https://doi.org/10.1007/978-3-030-86436-1_19

La conférence EC-TEL est classée A* par l'ATIEF, et classée B au classement *CORE*. En conséquence de cette distinction, nous avons été invités à soumettre une extension de papier à la revue *IEEE Transactions in Learning Technologies*. Le papier étendu a été publié en 2023.

Best Paper Award TICE 2014

SILVESTRE, F., VIDAL, P., & BROISIN, J. (2014a). Génération semi-automatique de tests d'auto-évaluation pourvus de feedback résultant de la prise de notes collaborative. *9eme Colloque Technologies de l'Information et de la Communication pour l'Enseignement (TICE 2014)*, 144-145

TICE 2014 fut la dernière édition de la conférence *Technologies de l'Information et de la Communication*. Elle n'apparaît donc plus dans aucun classement.

7.4 Diffusion du savoir, responsabilités et activités au sein des sociétés savantes ou associations

Participation comme expert à la conférence de Consensus du CNESCO

« Le Centre national d'étude des systèmes scolaires (Cnesco) analyse et accompagne des politiques, dispositifs et pratiques scolaires. Il vise à améliorer la connaissance des systèmes scolaires français et étrangers afin de créer des dynamiques de changement dans l'école, au profit de la réussite de tous les élèves. »

Le CNESCO organise chaque année une conférence de consensus autour d'un sujet imposé afin de regrouper experts et praticiens pour définir des recommandations concrètes pour améliorer les enseignements et apprentissages dans le système scolaire français. J'ai été invité à participer comme expert à la conférence de consensus 2022 sur le thème « L'évaluation en classe, au service de l'apprentissage des élèves »². Ma présentation portait sur l'usage du numérique au service de l'évaluation formative³.

Animateur de l'axe 1 du GTNum « Enseignement et apprentissage en situation(s) hybride(s) : vers une réinvention de la forme scolaire ? »

Les groupes thématiques numériques⁴ (GTNum), soutenus par la DNE pour deux ans à hauteur de 50K euros, produisent des travaux de recherche sur le numérique éducatif : pratiques pédagogiques ; technologie et impact ou potentiel pour l'enseignement ; enseignement, contexte et culture numérique. Ils sont animés par un ou plusieurs laboratoires accompagnés par les délégations académiques du numérique pour l'éducation (DANE). Ils visent à livrer des études, des enquêtes, des expériences / expérimentations et à documenter par des travaux de recherche des enjeux, des pratiques et des orientations stratégiques en matière de numérique éducatif. Les travaux résultants du groupe sont accessibles sur les *carnets d'hypothèses* édités en ligne par le Ministère de l'Éducation Nationale et de la Jeunesse⁵.

2. <https://www.cnesco.fr/evaluation-en-classe/>

3. https://www.youtube.com/watch?v=FMkVMb_1cT4

4. <https://edunumrech.hypotheses.org/>

5. <https://edunumrech.hypotheses.org/10324>

7.4. DIFFUSION DU SAVOIR, RESPONSABILITÉS ET ACTIVITÉS AU SEIN DES SOCIÉTÉS SAVANTES

- Période d'activité : Septembre 2020 à novembre 2023
- Taille du GTNum : 18 enseignants-chercheurs au total. L'axe 1 était composé de 5 enseignants-chercheurs de différentes disciplines et laboratoires (IRIT et LIRIS pour l'informatique, TECFA pour la psychologie cognitive).
- Entité porteuse : Laboratoire PERSEUS-Université de Lorraine
- Activités : animateur de l'axe 1 « Hybride distanciel/ présentiel »

Membre de l'ATIEF impliqué dans le groupe de travail « Épistémologie »

Je suis membre de l'Association des Technologies de l'Information pour l'Éducation et la Formation⁶ depuis 2014. Je suis engagé dans le groupe de travail « Épistémologie » depuis 2018 au travers de différentes activités dont les plus significatives sont :

- Membre du comité scientifique et participation à l'école d'automne « Méthodes de conduite de la recherche sur les environnements informatiques pour l'apprentissage humain (EIAH) »⁷ du 15 au 18 novembre 2022 à Leysin, Suisse.
- Organisation avec Nadine Mandran du symposium « La diversité de la recherche en EIAH par l'exemple »⁸ dans la cadre de la conférence EIAH 2023.

Diffusion 4C/ID

Le modèle 4C/ID (Four Components Instructional Design) a été développé dans les années 1990 par Van Merriënboer (1997) et ses collègues pour faciliter la conception d'enseignements ciblant l'apprentissage de réalisation de tâches complexes issues de situations professionnelles (VAN MERRI'ENBOER, 1997; VAN MERRIËNBOER & KIRSCHNER, 2017).

Au croisement de nos centres d'intérêt recherche et enseignement, 4C/ID influence à la fois mes travaux de recherche et ma manière d'enseigner. Plusieurs actions de diffusion de ce cadre de travail ont été menées allant de l'organisation de webinaires à la rédaction de chapitres de livres.

Webinaires pour la Centrale des IUT La Centrale des IUT⁹ est une association visant l'amélioration continue des enseignement et des apprentissages au sein des Instituts Universitaires de Technologie. J'ai créé et animé 2 webinaires sur 4C/ID¹⁰ pour la centrale des IUT en 2021. L'objectif de ces interventions était la présentation du cadre de

6. <http://www.atief.fr/>

7. <https://methodo-eiah.unige.ch/>

8. http://thedre.imag.fr/?page_id=1232

9. <https://www.centrale-iut.net/>

10. <https://www.centrale-iut.net/events/details/2675>

conception 4C/ID en illustrant sa mise en œuvre dans un enseignement du BUT Informatique.

Rédaction de chapitres livres J'ai contribué à 2 chapitres de livre à paraître fin 2023 dont un en auteur principal.

1. MOLINARI, G., NLEME ZE, Y. S., & SILVESTRE, F. (2023). Concevoir un cours en ligne, c'est du bricolage. In E. PAUKOVICS & E. SANCHEZ (Éd.), *Apprendre à distance* (Retz, collection Mythes et réalités, p. 63-76)
2. SILVESTRE, F., & PARMENTIER, J.-F. (2024). 4C/ID, une approche par compétences centrée sur les tâches authentiques. In A. MESSAOUI & C. PÉLISSIER (Éd.), *Vers l'approche par compétences - Théories et pratiques pour l'enseignement supérieur* (Presses des mines, collection Design Numérique)

Le livre aux éditions Retz est disponible en précommande¹¹.

7.5 Organisation colloques, conférences, journées d'étude

Organisation de la conférence ECTEL 2022

L'équipe TALENT de l'IRIT a organisé la 11ème *European Conference On Technology Enhanced Learning*¹². EC-TEL 2021 s'est déroulée à Toulouse du 12 au 15 septembre.

J'étais, entre autres, en charge de l'organisation de la participation des entreprises de la EdTech à l'évènement.

Organisation de la conférence EIAH 2013

L'équipe SIERA de l'IRIT a organisé la 6ème édition de la conférence *Environnements Informatiques pour l'Apprentissage Humain*. Je participais, en tant que doctorant, à la gestion de la logistique de la conférence.

7.6 Responsabilités scientifiques

Co-responsabilité du domaine d'application stratégique e-Education de l'IRIT

L'IRIT est structuré en 7 départements et en 5 domaines d'applications stratégiques (DAS). Le DAS e-Education vise à animer la communauté des chercheurs et chercheuses de l'IRIT travaillant sur des projets de recherche ayant pour champ d'application l'Éducation.

11. <https://www.editions-retz.com/enrichir-sa-pedagogie/mes-connaissances-educatives/apprendre-a-distance-9782725643960.html>

12. <https://ea-tel.eu/ectel2022>

- Période d'activité : Juin 2021 à aujourd'hui
- Co-responsable : Jean-Baptiste Raclet (IRIT, équipe ACADIE)
- Taille du DAS : 11 équipes de l'IRIT représentées par 9 professeurs, 21 MCF dont 3 HDR, 5 doctorants, 2 postdocs, 2 ingénieurs de recherche.
- Activités d'animation : Réunions mensuelles, communication via liste de diffusion, séminaire annuel incluant d'autres entités académiques ou économiques parties prenantes du secteur Éducation (laboratoires SHS, entreprises des EdTechs).

Contrats de recherche suite à appel à projet

B4MATIVE! 2019-2023 Projet financé sur appel à projet de la Direction du Numérique pour l'Éducation (DNE) issue du Ministère de l'Éducation Nationale de la Jeunesse et des Sports. Le projet mené en partenariat avec l'Académie de Nancy-Metz visait l'amélioration de la mise en œuvre de séquences d'évaluation formative en classe. La thèse de Rialy Andriamiseza ainsi qu'un ingénieur de recherche sur une période de 6 mois ont été financés par ce projet.

SCICONUM 2023-2026 Le projet SCICONUM a été lauréat de l'AMI « Compétences et Métiers d'Avenir » lancé dans le cadre plan d'investissement France 2030. Porté par l'entreprise Didask, le projet vise à développer les pratiques enseignantes fondées sur les résultats de recherche expérimentale d'une part, et l'amélioration des outils numériques mobilisables dans ces pratiques. Le projet finance ainsi un contrat postdoctoral sur 3 ans ainsi que de la prestation pour faire évoluer l'application Elaastic.

7.7 Liste des publications

Articles de journal

- ANDRIAMISEZA, R., SILVESTRE, F., PARMENTIER, J.-F., & BROISIN, J. (2023). How Learning Analytics Can Help Orchestration of Formative Assessment? Data-driven recommendations for Technology-Enhanced Learning. *IEEE Transactions on Learning Technologies*, 1-16. <https://doi.org/10.1109/tlt.2023.3265528>
- RACLET, J.-B., SILVESTRE, F., & PONS, M. (2021). Git4School : un tableau de bord pour assister la prise de décisions de l'enseignant lors des cours de génie logiciel. *Sciences et Technologies de l'Information et de la Communication pour l'Éducation et la Formation (STICEF)*, 28(3), 1-20.

SILVESTRE, F., VIDAL, P., & BROISIN, J. (2017). Un nouveau processus d'évaluation pour améliorer la qualité des feedbacks dans les tests en ligne. *Sciences et Technologies de l'Information et de la Communication pour l'Éducation et la Formation (STICEF)*, 24(1), 183-206.

Chapitres de livres

MOLINARI, G., NLEME ZE, Y. S., & SILVESTRE, F. (2023). Concevoir un cours en ligne, c'est du bricolage. In E. PAUKOVICS & E. SANCHEZ (Éd.), *Apprendre à distance* (Retz, collection Mythes et réalités, p. 63-76).

SILVESTRE, F., & PARMENTIER, J.-F. (2024). 4C/ID, une approche par compétences centrée sur les tâches authentiques. In A. MESSAOUI & C. PÉLISSIER (Éd.), *Vers l'approche par compétences - Théories et pratiques pour l'enseignement supérieur* (Presses des mines, collection Design Numérique).

Articles de conférences

ANDRIAMISEZA, R., SILVESTRE, F., PARMENTIER, J.-F., & BROISIN, J. (2021a). Data-informed Decision-making in TEFA Processes : An Empirical Study of a Process Derived from Peer-Instruction. *Proceedings of the Eighth ACM Conference on Learning @ Scale*. <https://doi.org/10.1145/3430895.3460153>

ANDRIAMISEZA, R., SILVESTRE, F., PARMENTIER, J.-F., & BROISIN, J. (2021b). Recommendations for Orchestration of Formative Assessment Sequences : A Data-Driven Approach. In *Technology-Enhanced Learning for a Free, Safe, and Sustainable World* (p. 245-259). Springer International Publishing. https://doi.org/10.1007/978-3-030-86436-1_19

ANDRIAMISEZA, R., SILVESTRE, F., PARMENTIER, J.-F., & BROISIN, J. (2021d). Vers la conception de feedback pour enseignants dans un contexte d'évaluation formative à grande échelle : une approche analytique. *10e Conférence sur les Environnements Informatiques pour l'Apprentissage Humain*, 46-57. <https://hal.archives-ouvertes.fr/hal-03292736>

OLLIVIER, D., SILVESTRE, F., RACLET, J.-B., LESCURE, E., & BROISIN, J. (2023a). Designing a Revision System : An Exploratory Qualitative Study to Identify the Needs of French Teachers and Students. In O. VIBERG, I. JIVET, P. J. MUÑOZ-MERINO, M. PERIFANOU & T. PAPATHOMA (Éd.), *Responsive and Sustainable Educational Futures* (p. 294-307). Springer Nature Switzerland.

OLLIVIER, D., SILVESTRE, F., RACLET, J.-B., LESCURE, E., & BROISIN, J. (2023b). Conception d'un système de révisions : une étude qualitative exploratoire pour identifier les besoins des enseignants et des élèves. In S. IKSAL, C. PÉLISSIER, J.-M. GILLIOT & P. MARZIN-JANVIER (Éd.), *11ème Conférence sur les Environnements*

- Informatiques pour l'Apprentissage Humain (EIAH 2023)* (p. 36-41). <https://hal.science/hal-04164903>
- PARMENTIER, J.-F., & SILVESTRE, F. (2019). La (dé-)synchronisation des transitions dans un processus d'évaluation formative exécuté à distance : impact sur l'engagement des étudiants. *9ème Conférence sur les Environnements Informatiques pour l'Apprentissage Humain (EIAH 2019)*, 97-108.
- PONS, M., BRUEL, J.-M., RACLET, J.-B., & SILVESTRE, F. (2023a). Finding behavioral indicators from contextualized commits in software engineering courses with process mining. *Frontiers In Software Engineering Education*, à paraître.
- PONS, M., BRUEL, J.-M., RACLET, J.-B., & SILVESTRE, F. (2023b). Traçabilité by design : conception d'un système interactif pour améliorer la génération automatique de traces Git pendant une activité d'apprentissage. *11ème Conférence sur les Environnements Informatiques pour l'Apprentissage Humain (EIAH 2023)*, 68-79.
- PONS, M., BRUEL, J.-M., RACLET, J.-B., & SILVESTRE, F. (2023c). Traceability by Design : Design of an Interactive System to Improve the Automatic Generation of Git Traces During a Learning Activity. *European Conference on Technology Enhanced Learning*, 611-617.
- RACLET, J.-B., & SILVESTRE, F. (2020). Git4School : A Dashboard for Supporting Teacher Interventions in Software Engineering Courses. In *Addressing Global Challenges and Quality Education* (p. 392-397). Springer International Publishing. https://doi.org/10.1007/978-3-030-57717-9_33
- RACLET, J.-B., SILVESTRE, F., & PONS, M. (2020). Mise en oeuvre d'approches pédagogiques fondées sur des pratiques de l'industrie du logiciel pour l'apprentissage de la programmation. *Didapro 8 - Colloque de didactique de l'informatique*.
- SILVESTRE, F., JOSEPH, M., DESSUS, P., PARMENTIER, J.-F., POUJET, L., REBAI, I., & TRANIER, J. (2016). L'informatique au service de la démocratisation des enseignements fondés sur l'évaluation formative. *Les Grands Challenges des ORPHEE RDV 2017*, 17-19.
- SILVESTRE, F., & RACLET, J.-B. (2018). Développement dirigé par les tests et revue de code par les pairs pour l'apprentissage de la programmation. *LudoviaCH*, 13-15.
- SILVESTRE, F., VIDAL, P., & BROISIN, J. (2014a). Génération semi-automatique de tests d'auto-évaluation pourvus de feedback résultant de la prise de notes collaborative. *9ème Colloque Technologies de l'Information et de la Communication pour l'Enseignement (TICE 2014)*, 144-145.
- SILVESTRE, F., VIDAL, P., & BROISIN, J. (2014d). Tsaap-Notes—An Open Micro-blogging Tool for Collaborative Notetaking during Face-to-Face Lectures. *Advanced Learning Technologies (ICALT), 2014 IEEE 14th International Conference on*, 39-43.

- SILVESTRE, F., VIDAL, P., & BROISIN, J. (2015a). Online Tests Based on Contributions Provided by Teachers and Students during Face to Face Lectures. *Advanced Learning Technologies (ICALT), 2015 IEEE 15th International Conference on*, 29-33.
- SILVESTRE, F., VIDAL, P., & BROISIN, J. (2015b). Qualification semi-automatique de contributions d'apprenants pour l'intégration de feedbacks de qualité dans les tests en ligne. *7ème Conférence sur les Environnements Informatiques pour l'Apprentissage Humain (EIAH 2015)*, 210-221.
- SILVESTRE, F., VIDAL, P., & BROISIN, J. (2015c). Reflexive learning, socio-cognitive conflict and peer-assessment to improve the quality of feedbacks in online tests. In *Design for Teaching and Learning in a Networked World* (p. 339-351). Springer.

Thèse

- SILVESTRE, F. (2015). *Conception et mise en oeuvre d'un système d'évaluation formative pour les cours en face à face dans l'enseignement supérieur*. Université de Toulouse, Université Toulouse III-Paul Sabatier.

Logiciel

- SILVESTRE, F., TRANIER, J., & ANDRIAMISEZA, R. (2012). *Elaastic*. <https://ut3-toulouseinp.hal.science/hal-03418475>

Chapitre 8

Activités d'enseignement

Contenu

8.1 Enseignements au département informatique de l'IUT de Rodez	105
8.2 Autres enseignements à l'Université Toulouse Capitole	106
8.3 Enseignements à l'Université Toulouse 3 Paul Sabatier	106
8.4 Autres enseignements	107
8.5 Mise en œuvre de pédagogies actives	107

J'ai commencé ma carrière professionnelle comme professeur certifié de mathématiques (obtention du CAPES de Mathématiques et titularisation en 1995). J'ai enseigné les mathématiques en collège et lycée entre 1994 et 1998. Durant mon service national en coopération en Côte d'Ivoire, mon intérêt pour la pédagogie et la didactique des mathématiques m'ont conduit à occuper la dernière année de mon séjour les fonctions de conseiller pédagogique.

Entre 2000 et 2006, après ma reprise d'études en informatique, j'ai dispensé des cours en MIAGE à l'Université de Haute Alsace (institution principale dans laquelle j'ai reçu ma formation en informatique) en tant que vacataire : j'ai créé en 2000 le cours relatif à la programmation d'applications distribuées à l'aide d'un *framework* Java. J'ai aussi dispensé durant 3 ans un cours similaire à l'IUT de Castres dans le cadre du DUT MMI. Ma fonction de dirigeant d'entreprise ne me laisse plus le temps de poursuivre mes enseignements entre 2006 et 2012.

Entre septembre 2012 et août 2018, occupant un poste d'enseignant-chercheur associé à mi-temps puis à temps complet, j'ai dispensé pendant 6 ans mes enseignements en Master Informatique spécialité Développement Logiciel. J'étais alors en charge de divers modules en lien avec le développement d'applications distribuées (enseignements sur Java EE) ou en lien avec la qualité du développement (gestion de configuration, intégration, vérification et validation du logiciel).

J'ai été recruté en septembre 2018 comme Maître de Conférences à l'IUT de Rodez. J'y

enseigne dans le département Informatique depuis ma prise de fonction. Je suis intervenu sur les formations DUT Informatique et Licence Pro et intervins depuis septembre 2022 sur le Bachelor Universitaire de Technologie (BUT) informatique issue de la réforme de la licence professionnelle.

J'enseigne dans les thématiques variées : bases de données, programmation Web, qualité du développement logiciel, IHM, correspondance Objet/Relationnel. Sur les 5 dernières années, j'ai pu maintenir des enseignements au niveau Master.

Les enseignants présentés plus en détail ci-dessous concernent la période 2012-2023.

8.1 Enseignements au département informatique de l'IUT de Rodez

L'IUT de Rodez est une composante de l'Université Toulouse Capitole.

Enseignements en BUT informatique

Les enseignements au titre du BUT 2^{ème} année sont dispensés en formation initiale depuis septembre 2022. Les enseignements au titre du BUT 3^{ème} année sont dispensés en formation en alternance depuis septembre 2023.

Niveau	Intitulé	Nature	Effectifs	HETD
L2	Analyse métier	CM/TP	54	34,5
L2	Développement Web Côté serveur	CM/TD/TP	54	22
L2	Développement d'une application (SAÉ*)	TD/TP	54	16,5
L2	Qualité de développement	CM/TP	40	45
L2	Dév. d'une application complexe (SAÉ)	TD	40	2,5
L3	Initiation au management d'une équipe	CM/TD	32	12,5
L3	Qualité de développement	CM/TD/TP	32	24,5
L3	Optimisation d'une application complexe (SAÉ)	TD	32	7,5
L3	Développement avancé (Jakarta EE)	CM/TD/TP	32	33,5
L3	Tutorat d'alternance		1	12,5
			Total annuel	211

* Une SAÉ désigne une Situation d'Apprentissage et d'Évaluation prenant la forme d'un projet.

Enseignements en Licence Professionnelle MMS

Les enseignements en Licence Professionnelle Multimedia-Mobilité-Sécurité (MMS) ont été dispensés en formation en alternance entre septembre 2018 et août 2023.

Niveau	Intitulé	Nature	Effectifs	HETD
L3	Interaction Humain Machine	CM/TD	12	44
L3	Professionnalisation (Git, tests)	CM/TD	12	18
L3	Persistance des données	CM/TD	12	32
L3	Tutorat d'alternance		1	12,5
			Total annuel	106,5

Enseignements en DUT informatique

Les enseignements au titre du DUT ont été dispensés en formation initiale entre septembre 2019 et août 2022.

Niveau	Intitulé	Nature	Effectifs	HETD
L1	Programmation et administration BD	CM/TD	64	30
L2	Méthodes de production d'application	CM/TD	48	73,5
L2	Développement Web côté serveur	CM/TD	48	30
L2	XML et Web Services	CM/TD	32	25,5
L2	Encadrement de projet	TD	16	14
L2	Encadrement de stage		4	23
			Total annuel	196,5

8.2 Autres enseignements à l'Université Toulouse Capitole

Les enseignements ont été dispensés dans l'UFR Informatique de l'Université Toulouse Capitole en master 1 mention MIAGe spécialité Ingénierie Métier durant l'année 2022-2023.

Niveau	Intitulé	Nature	Effectifs	HETD
M1	Conduite de projet informatique	CM/TD	96	28
			Total annuel	28

8.3 Enseignements à l'Université Toulouse 3 Paul Sabatier

Les enseignements ont été dispensés en Master mention Informatique spécialité Développement logiciel du département informatique de la Faculté des Sciences et de l'Ingénierie entre septembre 2012 et août 2018.

Niveau	Intitulé	Nature	Effectifs	HETD
M1	Développements collaboratifs, gestion de configuration	CM/TD/TP	50	39
M1	Développements d'applications réparties Java EE	CM/TP	50	33
M1	Méthodes agiles	CM/TD/TP	50	34
M1	Management de projets informatiques	CM/TD	50	34
M2	Développements back-end orientés Java EE	CM/TD/TP	50	25
M2	Intégration, vérification, validation, qualification	CM/TD/TP	36	30
M2	Méthodes agiles	CM/TD/TP	36	25
M2	Programmation concurrente et répartie	TP	36	14
			Total annuel	234

8.4 Autres enseignements

J'ai participé à la mise en place du DIU « Enseigner l'informatique au Lycée » s'adressant aux enseignants du Lycée engagés dans l'enseignement de la nouvelle spécialité Numérique et Sciences Informatiques (NSI).

J'étais alors en charge des modules « Paradigmes de programmation » et du module de sensibilisation à la didactique de l'informatique.

8.5 Mise en œuvre de pédagogies actives

Fortement intéressé par l'apprentissage et l'enseignement (directement en lien avec mes activités de recherche), je mets en œuvre les pédagogies dites actives, c'est-à-dire les procédés pédagogiques engageant fortement les étudiants sur le plan cognitif. J'utilise pour cela des méthodes et procédés issus de la recherche en sciences de l'éducation fondée sur les preuves. Parmi les cadres et approches mobilisés résident 4C/ID¹ ou encore le *Peer Instruction*². Dans le contexte de mes travaux de recherche, je développe l'application *Elaastic*^{3, 4} permettant d'orchestrer des séquences d'évaluation formative promouvant l'argumentation écrite et l'évaluation par les pairs. J'utilise cette application très régulièrement dans mes séquences de cours, TD ou TP et ai permis un usage plus généralisé au sein de l'IUT et de plusieurs institutions du supérieur en proposant des formations et de l'accompagnement sur l'outil. J'ai porté en 2019 le projet SRESRI « Pédagogies actives à l'ère du BYOD » afin d'équiper les grandes salles de l'IUT des infrastructures permettant l'utilisation d'applications en ligne comme *Elaastic* pendant les cours. Le projet a été accepté et entièrement financé par la région. Les salles sont opérationnelles depuis la rentrée 2021. *Elaastic* a fait l'objet de plusieurs publications scientifiques détaillées dans le volet recherche.

J'ai introduit à l'IUT de Rodez l'enseignement de la gestion de version de code source

1. <https://www.4cid.org/>

2. <https://ciel.unige.ch/2015/05/les-7-choses-que-vous-devriez-savoir-sur-linstruction-par-ses-pairs-peer-instruction/>

3. <https://www.irit.fr/elaastic/>

4. <https://sia.univ-toulouse.fr/initiatives-pedagogiques/plateforme-elaastic>

avec Git ainsi que son utilisation sur différents modules en vue de la remise des travaux d'étudiants et de leur suivi en utilisant l'application Git4School⁵. Git4School est un tableau de bord à destination des enseignants facilitant la prise de décision à partir de données objectives sur les contributions des étudiants au cours de leur apprentissage. Git4School a fait l'objet de plusieurs publications scientifiques détaillées dans le volet recherche.

5. <https://git4school.firebaseio.com/>

Bibliographie

- ANDERSON, T., & SHATTUCK, J. (2012). Design-based research : A decade of progress in education research? *Educational researcher*, 41(1), 16-25.
- ANDRIAMISEZA, R. (2022). *Learning Analytics-Based Formative Assessment Recommendations for Technology-Enhanced Learning Practices* (Theses 2022TOU30186). Université Paul Sabatier - Toulouse III. <https://theses.hal.science/tel-04006037>
- ANDRIAMISEZA, R., SILVESTRE, F., PARMENTIER, J.-F., & BROISIN, J. (2021a). Data-informed Decision-making in TEFA Processes : An Empirical Study of a Process Derived from Peer-Instruction. *Proceedings of the Eighth ACM Conference on Learning @ Scale*. <https://doi.org/10.1145/3430895.3460153>
- ANDRIAMISEZA, R., SILVESTRE, F., PARMENTIER, J.-F., & BROISIN, J. (2021b). Recommendations for Orchestration of Formative Assessment Sequences : A Data-Driven Approach. In *Technology-Enhanced Learning for a Free, Safe, and Sustainable World* (p. 245-259). Springer International Publishing. https://doi.org/10.1007/978-3-030-86436-1_19
- ANDRIAMISEZA, R., SILVESTRE, F., PARMENTIER, J.-F., & BROISIN, J. (2021c). Vers la conception de feedback pour enseignants dans un contexte d'évaluation formative à grande échelle : une approche analytique. *10e Conférence sur les Environnements Informatiques pour l'Apprentissage Humain*, 46-57. <https://hal.archives-ouvertes.fr/hal-03292736>
- ANDRIAMISEZA, R., SILVESTRE, F., PARMENTIER, J.-F., & BROISIN, J. (2023). How Learning Analytics Can Help Orchestration of Formative Assessment ? Data-driven recommendations for Technology-Enhanced Learning. *IEEE Transactions on Learning Technologies*, 1-16. <https://doi.org/10.1109/tlt.2023.3265528>
- APRIYANTI, D., & AYU, M. (2020). Think-Pair-Share : Engaging Students in Speaking Activities in Classroom. *Journal of English Language Teaching and Learning*, 1(1), 13-19.
- ATIEF. (2020). *Avis du Conseil d'Administration de l'ATIEF en réponse aux propositions du Conseil Scientifique de l'Éducation Nationale*. <http://www.atief.fr/content/avis-du-conseil-d%E2%80%99administration-de-l%E2%80%99atief-en-r%C3%A9ponse-aux-propositions-du-conseil>

- ATKINSON, R. K., DERRY, S. J., RENKL, A., & WORTHAM, D. (2000). Learning from examples : Instructional principles from the worked examples research. *Review of educational research*, 70(2), 181-214.
- BANGOR, A., KORTUM, P., & MILLER, J. (2009). Determining what individual SUS scores mean : Adding an adjective rating scale. *Journal of usability studies*, 4(3), 114-123.
- BIGGS, J. (1996). Enhancing teaching through constructive alignment. *Higher education*, 32(3), 347-364.
- BIGGS, J., TANG, C., & KENNEDY, G. (2022). *Ebook : Teaching for Quality Learning at University 5e*. McGraw-hill education (UK).
- BIRD, S. (2011). Effects of distributed practice on the acquisition of second language English syntax—ERRATUM. *Applied Psycholinguistics*, 32(2), 435-452.
- BLACK, P., & WILIAM, D. (1998). Assessment and Classroom Learning [Publisher : Routledge _eprint : <https://doi.org/10.1080/0969595980050102>]. *Assessment in Education : Principles, Policy & Practice*, 5(1), 7-74. <https://doi.org/10.1080/0969595980050102>
- BLACK, P., & WILIAM, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability(formerly : Journal of Personnel Evaluation in Education)*, 21(1), 5. <https://doi.org/10.1007/s11092-008-9068-5>
- BORREGO, M., CUTLER, S., FROYD, J., PRINCE, M., & HENDERSON, C. (2011). Faculty use of research based instructional strategies. *Australasian Association for Engineering Education Conference*, 448-453.
- BROISIN, J. (2020). *Towards intelligent technology-enhanced learning solutions for transforming higher education : contributions and future directions*. Université Toulouse 3 Paul Sabatier.
- BROOKE, J. (1996). SUS-A quick and dirty usability scale. *Usability evaluation in industry*, 189(194), 4-7.
- BUCHS, C., DARNON, C., QUIAMZADE, A., MUGNY, G., & BUTERA, F. (2008). Conflits et apprentissage. Régulation des conflits sociocognitifs et apprentissage. *Revue française de pédagogie. Recherches en éducation*, (163), 105-125.
- BULL, J., & MCKENNA, C. (2003). A blueprint for computer-assisted assessment.
- BURNSTEIN, R. A., & LEDERMAN, L. M. (2001). Using wireless keypads in lecture classes [Publisher : American Association of Physics Teachers]. *The Physics Teacher*, 39(1), 8-11. <https://doi.org/10.1119/1.1343420>
- CARNEGIE MELLON UNIVERSITY. (s. d.). Formative vs Summative Assessment. Récupérée 20 décembre 2023, à partir de <https://www.cmu.edu/teaching/assessment/basics/formative-summative.html>
- CARPENTER, S. K., CEPEDA, N. J., ROHRER, D., KANG, S. H., & PASHLER, H. (2012). Using spacing to enhance diverse forms of learning : Review of recent research and implications for instruction. *Educational Psychology Review*, 24(3), 369-378.

- CEPEDA, N. J., PASHLER, H., VUL, E., WIXTED, J. T., & ROHRER, D. (2006). Distributed practice in verbal recall tasks : A review and quantitative synthesis. *Psychological bulletin*, 132(3), 354-380.
- CHANOCK, K. (2000). Comments on essays : do students understand what tutors write? *Teaching in Higher Education*, 5.
- CHENG, W., SHEN, Y., HUANG, L., & ZHU, Y. (2019). Incorporating interpretability into latent factor models via fast influence analysis. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 885-893.
- CHI, M. T. H., ADAMS, J., BOGUSCH, E. B., BRUCHOK, C., KANG, S., LANCASTER, M., LEVY, R., LI, N., MCELDOON, K. L., STUMP, G. S., WYLIE, R., XU, D., & YAGHMOURIAN, D. L. (2018). Translating the ICAP Theory of Cognitive Engagement Into Practice. *Cognitive Science*, 42(6), 1777-1832. <https://doi.org/10.1111/cogs.12626>
- CHI, M. T. H., & 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>
- CHNG, L., & GURVITCH, R. (2018). Using Plickers as an Assessment Tool in Health and Physical Education Settings. *Journal of Physical Education, Recreation & Dance*, 89(2), 19-25. <https://doi.org/10.1080/07303084.2017.1404510>
- CLARIANA, R. B. (1993). A Review of Multiple-Try Feedback in Traditional and Computer-Based Instruction. *Journal of Computer-Based Instruction*, 20(3), 67-74.
- CLINE, K. S. (2006). Sharing Teaching Ideas : Classroom Voting in Mathematics [Publisher : National Council of Teachers of Mathematics Section : The Mathematics Teacher]. *The Mathematics Teacher*, 100(2), 100-104. <https://doi.org/10.5951/MT.100.2.0100>
- CLOW, D. (2012). The learning analytics cycle : closing the loop effectively. *Proceedings of the 2nd international conference on learning analytics and knowledge*, 134-138.
- CRESWELL, J. W., & CLARK, V. L. P. (2017). *Designing and conducting mixed methods research*. Sage publications.
- CRESWELL, J. W., FETTERS, M. D., & IVANKOVA, N. V. (2004). Designing a mixed methods study in primary care. *The Annals of Family Medicine*, 2(1), 7-12.
- CROUCH, C. H., & MAZUR, E. (2001). Peer Instruction : Ten years of experience and results [Publisher : American Association of Physics Teachers]. *American Journal of Physics*, 69(9), 970-977. <https://doi.org/10.1119/1.1374249>
- CSEN. (2020). *Recommandations pédagogiques pour accompagner le confinement et sa sortie*. <https://www.education.gouv.fr/csen-recommandations-pedagogiques-COVID19>
- CSIKSZENTMIHALYI, M. (2013). *Flow : The psychology of happiness*. Random House.

- DAMASIO, A. R. (2006). *L'erreur de Descartes : la raison des émotions*. Odile Jacob.
- DE CLERCQ, M., FRENAY, M., AZZI, A., KLEIN, O., & GALAND, B. (2021). All you need is self-determination : Investigation of PhD students' motivation profiles and their impact on the doctoral completion process. *International Journal of Doctoral Studies*, 16, 189.
- DEHAENE, S. (2013). Les quatre piliers de l'apprentissage, ou ce que nous disent les neurosciences. *Paris Tech Review*.
- DEMPSTER, F. N., & FARRIS, R. (1990). The spacing effect : research and practice. *Journal of Research & Development in Education*.
- DOCHY, F., SEGERS, M., & SLUIJSMANS, D. (1999). The use of self-, peer and co-assessment in higher education : A review. *Studies in Higher education*, 24(3), 331-350.
- DUFLO, E., & BANERJEE, A. V. (2019). *Repenser la pauvreté*. Média Diffusion.
- FAGEN, A. P., CROUCH, C. H., & MAZUR, E. (2002). Peer instruction : Results from a range of classrooms. *The physics teacher*, 40(4), 206-209.
- FALCONER, J. L. (2004). Use of ConcepTests and Instant Feedback in Thermodynamics [Number : 1]. *Chemical Engineering Education*, 38(1), 64-67. Récupérée 29 mars 2023, à partir de <https://journals.flvc.org/cee/article/view/122696>
- FITZGERALD, D. (2013). Employing think-pair-share in associate degree nursing curriculum. *Teaching and Learning in Nursing*, 8(3), 88-90.
- FREEMAN, M., & MCKENZIE, J. (2014). Aligning peer assessment with peer learning for large classes : the case for an online self and peer assessment system. In *Peer learning in higher education* (p. 156-169). Routledge.
- GAUCI, S. A., DANTAS, A. M., WILLIAMS, D. A., & KEMM, R. E. (2009). Promoting student-centered active learning in lectures with a personal response system [Publisher : American Physiological Society]. *Advances in Physiology Education*, 33(1), 60-71. <https://doi.org/10.1152/advan.00109.2007>
- GUO, P. J., KIM, J., & RUBIN, R. (2014). How Video Production Affects Student Engagement : An Empirical Study of MOOC Videos. *Proceedings of the First ACM Conference on Learning @ Scale Conference*, 41-50. <https://doi.org/10.1145/2556325.2566239>
- HATTIE, J., & TIMPERLEY, H. (2007). The Power of Feedback [Publisher : American Educational Research Association]. *Review of Educational Research*, 77(1), 81-112. <https://doi.org/10.3102/003465430298487>
- HIGGINS, R. (2000a). *Be more critical : rethinking assessment feedback*. DYE, V, PC4000.
- HIGGINS, R. (2000b). "Be more critical!" : Rethinking assessment feedback [Publisher : Education-line]. Récupérée 18 février 2021, à partir de <http://www.leeds.ac.uk/educol/documents/00001548.htm>
- HOUDÉ, O. (2019a). *Apprendre à résister*. Le Pommier. <https://hal.science/hal-04047717>

- HOUDÉ, O. (2019b). *Apprendre à résister*. le Pommier.
- KADDOURA, M. (2013). Think pair share : A teaching learning strategy to enhance students' critical thinking. *Educational Research Quarterly*, 36(4), 3-24.
- KAHNEMAN, D. (2016). *Système 1 / Système 2 : Les deux vitesses de la pensée*. Clé des champs.
- KAHNEMAN, D., SIBONY, O., & SUNSTEIN, C. R. (2021). *Noise : Pourquoi nous faisons des erreurs de jugement et comment les éviter*. Odile Jacob.
- KAY, R. H., & LESAGE, A. (2009). Examining the benefits and challenges of using audience response systems : A review of the literature. *Computers & Education*, 53(3), 819-827. <https://doi.org/10.1016/j.compedu.2009.05.001>
- KOTHIYAL, A., MURTHY, S., & IYER, S. (2014). Think-pair-share in a large CS1 class : Does learning really happen ? *Proceedings of the 2014 conference on Innovation & technology in computer science education*, 51-56.
- KUHN, T. (2008). *La structure des révolutions scientifiques*. Flammarion, coll. « Champs / 791 », 1ère édition en 1962.
- LAKATOS, I. (1994). *Histoire et méthodologie des sciences : programme de recherche et reconstruction rationnelle*. PUF.
- LATIMIER, A. (2019). *Optimisation de l'apprentissage par récupération en mémoire pour promouvoir la rétention à long terme de nouvelles connaissances* (thèse de doct.). Université Paris sciences et lettres.
- LATOUR, B. (2001). *Le métier de chercheur : regard d'un anthropologue*. Editions Quae.
- LE MONDE. (2023). *L'enseignement supérieur face à une baisse du nombre d'inscrits, sauf dans le privé*. Récupérée 7 septembre 2023, à partir de https://www.lemonde.fr/societe/article/2023/07/29/enseignement-superieur-une-baisse-du-nombre-d-inscrits-sauf-dans-le-privé_6183813_3224.html#xtor=AL-32280270-%5Bmail%5D-%5Bios%5D
- LEA, M. (2001). Computer conferencing and assessment : new ways of writing in higher education. *Studies in Higher Education*, 26(2), 163-181.
- LEDoux, J. (2005). *Le cerveau des émotions*. Odile Jacob.
- LIGHTNER, J., & TOMASWICK, L. (2017). *Active learning—think, pair, share*. Récupérée 15 août 2023, à partir de <https://www.kent.edu/ctl/teaching-tools-think-pair-share>
- LOPEZ, L. M. (2017). *La régulation des apprentissages en classe : Guide pédagogique*. De Boeck (Pédagogie et Formation).
- LYMAN, F. T. (1981). The responsive classroom discussion : The inclusion of all students. A. Anderson (Ed.), *Mainstreaming Digest, College Park : University of Maryland Press*, 109, 113.
- MAZUR, E. (1997). Peer instruction : Getting students to think in class [Publisher : American Institute of Physics]. *AIP Conference Proceedings*, 399(1), 981-988. <https://doi.org/10.1063/1.53199>

- McCONNELL, D. A., STEER, D. N., OWENS, K. D., KNOTT, J. R., VAN HORN, S., BOROWSKI, W., DICK, J., FOOS, A., MALONE, M., MCGREW, H., GREER, L., & HEANEY, P. J. (2006). Using Conceptests to Assess and Improve Student Conceptual Understanding in Introductory Geoscience Courses [Publisher : Routledge _eprint : <https://doi.org/10.5408/1089-9995-54.1.61>]. *Journal of Geoscience Education*, 54(1), 61-68. <https://doi.org/10.5408/1089-9995-54.1.61>
- MCMILLAN, J. H., & HEARN, J. (2008). Student self-assessment : The key to stronger student motivation and higher achievement. *Educational horizons*, 87(1), 40-49.
- MILLER, T. (2009). Formative computer-based assessment in higher education : the effectiveness of feedback in supporting student learning [Publisher : Routledge]. *Assessment & Evaluation in Higher Education*, 34(2), 181-192. <https://doi.org/10.1080/02602930801956075>
- MILLER, T. (2019). Explanation in artificial intelligence : Insights from the social sciences. *Artificial intelligence*, 267, 1-38.
- MOLINARI, G., NLEME ZE, Y. S., & SILVESTRE, F. (2023). Concevoir un cours en ligne, c'est du bricolage. In E. PAUKOVICS & E. SANCHEZ (Éd.), *Apprendre à distance* (Retz, collection Mythes et réalités, p. 63-76).
- NICOL, D. J., & BOYLE, J. T. (2003). Peer instruction versus class-wide discussion in large classes : a comparison of two interaction methods in the wired classroom. *Studies in Higher Education*, 28(4), 457-473.
- NICOL, D. J., & MACFARLANE-DICK, D. (2006). Formative assessment and self-regulated learning : a model and seven principles of good feedback practice. *Studies in higher education*, 31(2), 199-218.
- NICOL, D. J., & MACFARLANE-DICK, D. (2006). Formative assessment and self-regulated learning : a model and seven principles of good feedback practice [Publisher : Routledge]. *Studies in Higher Education*, 31(2), 199-218. <https://doi.org/10.1080/03075070600572090>
- OCDE. (2022). PISA 2015 - PISA. Récupérée 20 décembre 2023, à partir de <https://www.oecd.org/pisa/test/>
- OLLIVIER, D., SILVESTRE, F., RACLET, J.-B., LESCURE, E., & BROISIN, J. (2023a). Designing a Revision System : An Exploratory Qualitative Study to Identify the Needs of French Teachers and Students. In O. VIBERG, I. JIVET, P. J. MUÑOZ-MERINO, M. PERIFANOU & T. PAPATHOMA (Éd.), *Responsive and Sustainable Educational Futures* (p. 294-307). Springer Nature Switzerland.
- OLLIVIER, D., SILVESTRE, F., RACLET, J.-B., LESCURE, E., & BROISIN, J. (2023b). Conception d'un système de révisions : une étude qualitative exploratoire pour identifier les besoins des enseignants et des élèves. In S. IKSAL, C. PÉLISSIER, J.-M. GILLIOT & P. MARZIN-JANVIER (Éd.), *11ème Conférence sur les Environnements*

- Informatiques pour l'Apprentissage Humain (EIAH 2023)* (p. 36-41). <https://hal.science/hal-04164903>
- PARMENTIER, J.-F. (2018). How to quantify the efficiency of a pedagogical intervention with a single question. *Physical Review Physics Education Research*, 14(2), 020116.
- PARMENTIER, J.-F., LAMINE, B., & BONNAFÉ, S. (2015). Changer les conceptions en mécanique des étudiants en L1 à l'Université. In A. F. d. MÉCANIQUE (Éd.), *CFM 2015 - 22ème Congrès Français de Mécanique*. AFM, Maison de la Mécanique, 39/41 rue Louis Blanc - 92400 Courbevoie. Récupérée 22 juillet 2022, à partir de <https://hal.archives-ouvertes.fr/hal-03444641>
- PARMENTIER, J.-F., & SILVESTRE, F. (2019). La (dé-)synchronisation des transitions dans un processus d'évaluation formative exécuté à distance : impact sur l'engagement des étudiants. *9ème Conférence sur les Environnements Informatiques pour l'Apprentissage Humain (EIAH 2019)*, 97-108.
- PERRENOUD, P. (1998). From Formative Evaluation to a Controlled Regulation of Learning Processes. Towards a wider conceptual field [Publisher : Routledge _eprint : <https://doi.org/10.1080/0969595980050105>]. *Assessment in Education : Principles, Policy & Practice*, 5(1), 85-102. <https://doi.org/10.1080/0969595980050105>
- PHELPS, R. P. (2012). The effect of testing on student achievement, 1910–2010. *International Journal of Testing*, 12(1), 21-43.
- POPPER, K. (1973). *La logique de la découverte scientifique*. Payot.
- QUIN, F., WEYNS, D., GALSTER, M., & SILVA, C. C. (2023). A/b testing : A systematic literature review. *arXiv preprint arXiv :2308.04929*.
- RABARDEL, P. (1995). *Les hommes et les technologies ; approche cognitive des instruments contemporains*. Armand Colin. <https://hal.science/hal-01017462>
- RACLET, J.-B., & SILVESTRE, F. (2020). Git4School : A Dashboard for Supporting Teacher Interventions in Software Engineering Courses. In *Addressing Global Challenges and Quality Education* (p. 392-397). Springer International Publishing. https://doi.org/10.1007/978-3-030-57717-9_33
- RATMININGSIH, N. M., ARTINI, L. P., & PADMADEWI, N. N. (2017). Incorporating self and peer assessment in reflective teaching practices. *International Journal of Instruction*, 10(4), 165-184.
- RENKL, A. (2014). Toward an instructionally oriented theory of example-based learning. *Cognitive science*, 38(1), 1-37.
- RIBERA, M., & LAPEDRIZA GARCIA, A. (2019). Can we do better explanations? A proposal of user-centered explainable AI.
- RICKETTS, C., & WILKS, S. J. (2002). Improving Student Performance Through Computer-based Assessment : Insights from recent research [Publisher : Routledge]. *Assessment & Evaluation in Higher Education*, 27(5), 475-479. <https://doi.org/10.1080/0260293022000009348>

- RIES, E. (2012). *Lean startup : adoptez l'innovation continue*. Pearson.
- ROEDIGER, H. L. I., & KARPICKE, J. D. (2006). Test-Enhanced Learning : Taking Memory Tests Improves Long-Term Retention [PMID : 16507066]. *Psychological Science*, *17*(3), 249-255. <https://doi.org/10.1111/j.1467-9280.2006.01693.x>
- ROHRER, D., & TAYLOR, K. (2006). The effects of overlearning and distributed practise on the retention of mathematics knowledge. *Applied Cognitive Psychology : The Official Journal of the Society for Applied Research in Memory and Cognition*, *20*(9), 1209-1224.
- ROLHEISER, C., & ROSS, J. A. (2001). Student self-evaluation : What research says and what practice shows. *Plain talk about kids*, *43*(57).
- ROWLAND, C. A. (2014). The effect of testing versus restudy on retention : a meta-analytic review of the testing effect. *Psychological bulletin*, *140*(6), 1432.
- RUDOLPH, A. L., LAMINE, B., JOYCE, M., VIGNOLLES, H., & CONSIGLIO, D. (2014). Introduction of interactive learning into French university physics classrooms [Publisher : American Physical Society]. *Physical Review Special Topics - Physics Education Research*, *10*(1), 010103. <https://doi.org/10.1103/PhysRevSTPER.10.010103>
- SADLER, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, *18*(2), 119-144. <https://doi.org/10.1007/BF00117714>
- SCHAFFER, J. B., KONSTAN, J., & RIEDL, J. (1999). Recommender systems in e-commerce. *Proceedings of the 1st ACM conference on Electronic commerce*, 158-166.
- SCHLATTER, M. D. (2002). Writing Conceptests for a Multivariable Calculus Class [Publisher : Taylor & Francis _eprint : <https://doi.org/10.1080/10511970208984036>]. *PRIMUS*, *12*(4), 305-314. <https://doi.org/10.1080/10511970208984036>
- SCOULLER, K. (1998). The influence of assessment method on students' learning approaches : Multiple choice question examination versus assignment essay. *Higher education*, *35*(4), 453-472.
- SEIBERT HANSON, A. E., & BROWN, C. M. (2020). Enhancing L2 learning through a mobile assisted spaced-repetition tool : an effective but bitter pill ? *Computer Assisted Language Learning*, *33*(1-2), 133-155.
- SHAFFER, D. M., & COLLURA, M. J. (2009). Evaluating the Effectiveness of a Personal Response System in the Classroom. *Teaching of Psychology*, *36*(4), 273-277. <https://doi.org/10.1080/00986280903175749>
- SILVESTRE, F. (2015). *Conception et mise en oeuvre d'un système d'évaluation formative pour les cours en face à face dans l'enseignement supérieur*. Université de Toulouse, Université Toulouse III-Paul Sabatier.
- SILVESTRE, F. (2019). Engagement cognitif et activités pédagogiques éventuellement informatisées. *Apprentissage du management et neuropédagogie, Journées de pratique et de recherche, TSM*. <https://hal.science/hal-04181170>

- SILVESTRE, F. (2022). DRAXO, Un cadre pour l'évaluation formative par les pairs. *Quatrième Journées des INSPE d'Occitanie*. <https://hal.science/hal-04186725>
- SILVESTRE, F. (2023). Évaluer c'est former! Quel apport du numérique? *Journée Pédagogie de l'ARIUT - Au lycée comme à l'IUT : Évaluer, c'est former*. <https://hal.science/hal-04185858>
- SILVESTRE, F., & PARMENTIER, J.-F. (2021). *Technology-Enhanced Formative Assessment : Does Changing the Way Students Confront Their Views Change Their Behaviour and Outcomes ?* [working paper or preprint]. <https://hal.science/hal-04179720>
- SILVESTRE, F., & PARMENTIER, J.-F. (2024). 4C/ID, une approche par compétences centrée sur les tâches authentiques. In A. MESSAOUI & C. PÉLISSIER (Éd.), *Vers l'approche par compétences - Théories et pratiques pour l'enseignement supérieur* (Presses des mines, collection Design Numérique).
- SILVESTRE, F., TRANIER, J., & ANDRIAMISEZA, R. (2012). *Elaastic*. <https://ut3-toulouseinp.hal.science/hal-03418475>
- SILVESTRE, F., VIDAL, P., & BROISIN, J. (2014a). Génération semi-automatique de tests d'auto-évaluation pourvus de feedback résultant de la prise de notes collaborative. *9eme Colloque Technologies de l'Information et de la Communication pour l'Enseignement (TICE 2014)*, 144-145.
- SILVESTRE, F., VIDAL, P., & BROISIN, J. (2014b). Tsaap-Notes—An Open Micro-blogging Tool for Collaborative Notetaking during Face-to-Face Lectures. *Advanced Learning Technologies (ICALT), 2014 IEEE 14th International Conference on*, 39-43.
- SILVESTRE, F., VIDAL, P., & BROISIN, J. (2015a). Online Tests Based on Contributions Provided by Teachers and Students during Face to Face Lectures. *Advanced Learning Technologies (ICALT), 2015 IEEE 15th International Conference on*, 29-33.
- SILVESTRE, F., VIDAL, P., & BROISIN, J. (2015b). Qualification semi-automatique de contributions d'apprenants pour l'intégration de feedbacks de qualité dans les tests en ligne. *7ème Conférence sur les Environnements Informatiques pour l'Apprentissage Humain (EIAH 2015)*, 210-221.
- SILVESTRE, F., VIDAL, P., & BROISIN, J. (2015c). Reflexive learning, socio-cognitive conflict and peer-assessment to improve the quality of feedbacks in online tests. In *Design for Teaching and Learning in a Networked World* (p. 339-351). Springer.
- SILVESTRE, F., VIDAL, P., & BROISIN, J. (2017). Un nouveau processus d'évaluation pour améliorer la qualité des feedbacks dans les tests en ligne. *Sciences et Technologies de l'Information et de la Communication pour l'Éducation et la Formation (STICEF)*, 24(1), 183-206.
- SMITH, M. K., WOOD, W. B., KRAUTER, K., & KNIGHT, J. K. (2011). Combining peer discussion with instructor explanation increases student learning from in-class concept questions. *CBE—Life Sciences Education*, 10(1), 55-63.

- SOLARIS, S. f. L. A. R. (2012). *What is Learning Analytics ?* Récupérée 17 septembre 2023, à partir de <https://www.solaresearch.org/about/what-is-learning-analytics/>
- SWELLER, J. (2006). The worked example effect and human cognition. *Learning and instruction*.
- SWELLER, J. (2011). Cognitive load theory. In *Psychology of learning and motivation* (p. 37-76). Elsevier.
- TCHOUNIKINE, P. (2009). Précis de recherche en Ingénierie des EIAH [109 pages]. <https://hal.science/hal-00413694>
- THALER, R. H., & SUNSTEIN, C. R. (2022). *Nudge : la méthode douce pour inspirer la bonne décision*. Vuibert.
- TINTAREV, N., & MASTHOFF, J. (2015). Explaining recommendations : Design and evaluation. In *Recommender systems handbook* (p. 353-382). Springer.
- TOPPING, K. (1998). Peer assessment between students in colleges and universities. *Review of educational Research*, 68(3), 249-276.
- TOPPING, K. (2018). *Using peer assessment to inspire reflection and learning*. Routledge.
- TOPPING, K. J. (2009). Peer Assessment. *Theory Into Practice*, 48(1), 20-27. <https://doi.org/10.1080/00405840802577569>
- TVERSKY, A., & KAHNEMAN, D. (1974). Judgment under Uncertainty : Heuristics and Biases : Biases in judgments reveal some heuristics of thinking under uncertainty. *science*, 185(4157), 1124-1131.
- UHARI, M., RENKO, M., & SOINI, H. (2003). Experiences of using an interactive audience response system in lectures. *BMC Medical Education*, 3(1), 12. <https://doi.org/10.1186/1472-6920-3-12>
- UNESCO. (2022). *Les chiffres de l'enseignement supérieur en un coup d'œil*. Récupérée 7 août 2023, à partir de http://uis.unesco.org/sites/default/files/documents/f_unesco1015_brochure_web_fr.pdf
- VAN MERRI'ENBOER, J. J. (1997). *Training complex cognitive skills : A four-component instructional design model for technical training*. Educational Technology.
- VAN MERRI'ENBOER, J. J., & KIRSCHNER, P. A. (2017). *Ten steps to complex learning : A systematic approach to four-component instructional design*. Routledge.
- VENKATESH, V., THONG, J. Y., & XU, X. (2012). Consumer acceptance and use of information technology : extending the unified theory of acceptance and use of technology. *MIS quarterly*, 157-178.
- VICKREY, T., ROSPLOCH, K., RAHMANIAN, R., PILARZ, M., & STAINS, M. (2015). Research-Based Implementation of Peer Instruction : A Literature Review [Publisher : American Society for Cell Biology (Ise)]. *CBE—Life Sciences Education*, 14(1), es3. <https://doi.org/10.1187/cbe.14-11-0198>

- WANG, A. I., & TAHIR, R. (2020). The effect of using Kahoot! for learning – A literature review. *Computers & Education*, 149, 103818. <https://doi.org/10.1016/j.compedu.2020.103818>
- WASH, P. D. (2014). Taking advantage of mobile devices : Using Socrative in the classroom. *Journal of Teaching and Learning with Technology*, 99-101. <https://doi.org/10.14434/jotlt.v3n1.5016>
- WILIAM, D. (2007). Keeping learning on track : classroom assessment and the regulation of learning. In F. K. LESTER (Éd.), *In : Lester, FK, (ed.) Second Handbook of Research on Mathematics Teaching and Learning. (pp. 1053-1098). Information Age Publishing : Greenwich, Connecticut, USA. (2007) (p. 1053-1098). Information Age Publishing*. Récupérée 22 juillet 2022, à partir de <https://discovery.ucl.ac.uk/id/eprint/1507207/>
- ZAKRZEWSKI, S., & BULL, J. (1998). The Mass Implementation and Evaluation of Computer-based Assessments [Publisher : Routledge _eprint : <https://doi.org/10.1080/0260293980230203>]. *Assessment & Evaluation in Higher Education*, 23(2), 141-152. <https://doi.org/10.1080/0260293980230203>
- ZAROMB, F. M., & ROEDIGER, H. L. (2010). The testing effect in free recall is associated with enhanced organizational processes. *Memory & cognition*, 38(8), 995-1008.