



HAL
open science

Conception et mise en oeuvre d'un système d'évaluation formative pour les cours en face à face dans l'enseignement supérieur

Franck Silvestre

► To cite this version:

Franck Silvestre. Conception et mise en oeuvre d'un système d'évaluation formative pour les cours en face à face dans l'enseignement supérieur. Environnements Informatiques pour l'Apprentissage Humain. Université Toulouse III Paul Sabatier, 2015. Français. NNT : . tel-03296987

HAL Id: tel-03296987

<https://ut3-toulouseinp.hal.science/tel-03296987v1>

Submitted on 23 Jul 2021

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.



Université
de Toulouse

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par : *l'Université Toulouse 3 Paul Sabatier (UT3 Paul Sabatier)*

Présentée et soutenue le *25/11/2015* par :

Franck SILVESTRE

**Conception et mise en oeuvre d'un système d'évaluation formative pour
les cours en face à face dans l'enseignement supérieur**

JURY

Pierre TCHOUNIKINE	Professeur des Universités, Université Grenoble 1 Joseph Fourier	Président
Serge GARLATTI	Professeur des Universités, École TELECOM Bretagne	Rapporteur
Sébastien GEORGE	Professeur des Universités, Université du Maine	Rapporteur
Thierry NODENOT	Professeur des Universités, Université de Pau et des Pays de l'Adour	Examineur
Philippe VIDAL	Professeur des Universités, Université Toulouse 3 Paul Sabatier	Directeur de thèse
Julien BROISIN	Maître de conférences, Université Toulouse 3 Paul Sabatier	Encadrant de thèse

École doctorale et spécialité :

MITT : Domaine STIC : Réseaux, Télécoms, Systèmes et Architecture

Unité de Recherche :

Institut de Recherche en Informatique de Toulouse (UMR 5505)

Directeur(s) de Thèse :

Philippe VIDAL et Julien BROISIN

Rapporteurs :

Serge GARLATTI et Sébastien GEORGE

Remerciements

Remercier, c'est ajouter de la joie à la joie.

André Comte-Sponville

Il y a d'abord eu les échanges avec André Tricot, qui en raison de mon profil et de mes objectifs, m'a conseillé de rencontrer Philippe Vidal et Julien Broisin. Merci donc à André pour sa disponibilité, son écoute et ses conseils avisés. Merci à Philippe, à Julien ainsi qu'à Abdelmalek Benzekri d'avoir accueilli mon projet de thèse et de l'avoir soutenu jusqu'à son terme.

Merci à Julien pour son accompagnement, pour les discussions et les échanges, pour avoir su gérer avec tact l'encadrement d'un thésard "vintage" ; merci pour les soirées partagées en conférence qui ont permis de tisser d'autres liens.

Merci à Philippe Palanque pour sa disponibilité, sa bienveillance et pour le relai de bon nombre de publications relatives aux secteurs de l'éducation et des IHM.

Merci à Michelle Sibilla pour ses analyses sur le travail effectué : autant de retours précieux pour finir avec la distance nécessaire.

Merci à Rémi Venant, compagnon de bureau avec qui nous avons partagé idées et bons moments en conférence.

Merci à tous les membres de l'équipe SIERA et de l'équipe ICS pour tous les moments de convivialité partagés.

Ensuite, il y a eu les rencontres lors des colloques. Des rencontres scientifiques et humaines. Merci à Gaëlle Molinari pour les échanges dynamiques et enthousiastes, merci à Pierre-André Caron pour sa bienveillance et ses conseils pragmatiques. Merci à tous ceux qui ont bien voulu porter attention à mes travaux, qui ont réagi et partagé leur point de vue.

Je me rappelle la discussion avec Pierre Tchounikine à EIAH 2013, au début de la thèse, les échanges avec Sébastien George suite aux premières publications, les discussions sur l'agilité au service des projets de recherche avec Serge Garlatti ou encore la rencontre avec Thierry Nodenot lors d'une réunion du réseau Orphée. Ces quatre personnes ont accepté d'être membres du jury de ma thèse. Merci à Serge et Sébastien d'avoir accepté le rôle de rapporteur, merci pour leur investissement et leurs éclairages utiles pour les travaux futurs. Merci à Pierre et à Thierry d'avoir

accepté le rôle d'examineur ; le travail réalisé au cours de la thèse et celui à venir profiteront ainsi de toute la valeur apportée par leur expérience.

Ce travail n'aurait pas été le même sans le soutien de l'équipe pilotant le Master Développement Logiciel. Merci à Bernard Cherbonneau et à Henri Massié pour la confiance qu'ils m'ont accordée il y a maintenant quatre ans. Merci à Guillaume Cabanac sans qui la rencontre n'aurait pas eu lieu. Merci à Jean-Paul Arcangeli, Christine Maurel, Frederic Migeon, Jean-Paul Bodeveix et Jean-Baptiste Racllet pour la belle équipe que nous formons.

Fin 2014, la possibilité de passer à temps plein sur mon poste à l'Université s'est transformée en réalité. Ce passage à temps plein a joué un rôle décisif pour terminer la thèse dans des délais raisonnables. Je remercie Jean-Paul Arcangeli qui a porté et piloté le montage du dossier. Merci à tous ceux qui ont apporté leur appui pour qu'il aboutisse : Alain Crouzil, Jean-François Mazoin, Philippe Palanque et Abdelmalek Benzekri.

Les résultats des recherches menées pendant ces trois années intéressent aujourd'hui l'Université Toulouse 3 qui a déjà investi sur la mise en oeuvre de Tsaap-Notes. Ceci n'aurait pas été possible sans la rencontre avec Jean-François Parmentier et Brahime Lamine. Merci à eux pour l'élan et la visibilité locale qu'ils ont donnés aux travaux réalisés.

Sur le chemin de la thèse il y a aussi les amis d'avant la thèse : attentifs et curieux, encourageants, et toujours bienveillants. Merci à Anne et Nanou, à Christine et Laurent, à Fleur et Bastien. Merci à Matthieu et à Christophe, mes compagnons de séances ciné "vide cervelle". Merci à Laura, Émilie et Fred pour les joies (et les moments de tracs) partagés avec les Pink's !

Un merci tout particulier à John : frère d'arme sur bien des batailles. Vive nos résonances de raisons !

Merci à John, Nicolas et Sylvie qui ont eu la patience de relire, de commenter et de corriger le manuscrit.

Il y a bien sûr ceux qui sont là depuis toujours. Maman, Papa, Caroline, merci pour le regard et l'attention portés sur moi depuis aussi longtemps que je me souviens :

quelle meilleure source pour appréhender le monde et atteindre ce qui initialement pouvait paraître inaccessible ?

Merci à Danièle, Annie et Bernard : avec vous, "belle-famille" n'est pas une vaine expression. Merci à Véronique pour toutes les attentions.

Puis il y a mes deux soleils, mon épouse Marylène et mon fils Thomas. Dans les moments difficiles, dans les moments de faiblesse et de doute, ils sont les moyens et la fin. Merci pour la joie au quotidien.

Table des matières

Table des matières	III
Table des figures	VII
Liste des tableaux	X
Introduction	1
1 Introduction	3
1.1 Cadre scientifique	3
1.2 Contexte	5
1.3 Problématique de recherche et contributions	8
1.4 Organisation du mémoire	11
I État de l’art	13
2 L’évaluation formative	15
2.1 Introduction	15
2.2 Évaluation formative vs Évaluation sommative	16
2.3 Le feedback	17
2.4 L’auto-régulation	19
2.5 Auto-évaluation et évaluation par les pairs	21
2.6 Synthèse : éléments de la théorie de l’évaluation formative	23
2.7 Conclusion	26
3 Technologies et évaluation formative dans les amphithéâtres	29
3.1 Introduction	30
3.2 La méthode d’analyse	30
3.3 Les Systèmes de votes interactifs	37

3.4	Les systèmes d'échanges et de partages de textes	44
3.5	Synthèse de l'étude	48
3.6	Les verrous identifiés	50
3.7	Conclusion	53
II	Système d'évaluation d'audience	55
4	Modélisation standardisée de processus d'évaluation d'audience	57
4.1	Introduction	58
4.2	Définitions	58
4.3	Besoin d'un langage commun pour la description des processus	60
4.4	BPMN, une notation standard pour la description formelle de processus	61
4.5	Modèles de processus caractérisant un cours en face-à-face	65
4.6	Modèles de processus d'évaluation d'audience supportés par les SVI .	67
4.7	Le processus d'évaluation d'audience en N phases	72
4.8	Conclusion	74
5	Conception et mise en œuvre du SEA Tsaap-Notes	77
5.1	Introduction	78
5.2	Choix d'architecture technique	78
5.3	Itération 1 - SVI, micro-blogging et prise de notes collaborative : Tsaap-Notes	87
5.4	Itération 2 - La génération semi-automatique de tests de révisions . .	95
5.5	Itération 3 - Le SEA Tsaap	100
5.6	Conclusion	107
III	Expérimentations	111
6	Contexte général des expérimentations	113
6.1	Contexte humain	113
6.2	Contexte technique	114
6.3	Validité des expérimentations	114
7	Les résultats d'expérimentations	117
7.1	Expérimentation suite à l'itération 1	118

7.2	Expérimentation suite à l'itération 2	121
7.3	Expérimentation suite à l'itération 3	124
7.4	Conclusion	129
Conclusion et perspectives de recherches		135
8	Conclusion et perspectives de recherche	135
8.1	Les contributions	135
8.2	Perspectives de recherche	138
Bibliographie		139

Table des figures

1	Synthèse des résultats obtenus par Bloom	7
2	Modèle d'apprentissage auto-régulé (Nicol et Macfarlane-Dick, 2006) .	20
3	Le modèle CATAALYST de Roschelle <i>et al.</i> (2004)	59
4	Exemple de Processus privé	63
5	Exemple de Collaboration	64
6	Modèle du processus Cours Magistral	67
7	Modèle du processus Cours Interactif	68
8	Modèle du processus d'évaluation d'audience nominale	70
9	Modèle du processus d'évaluation d'audience "Instruction par les pairs"	71
10	Modèle du processus de soumission de réponses en N phases	73
11	La plateforme Java EE	81
12	Le patron de conception MVC	83
13	MVC - le découpage du modèle	83
14	MVC - la correspondance objet relationnel	83
15	L'Injection de dépendances	85
16	Impact de Tsaap-Notes sur le processus de Cours Interactif	89
17	Impact de Tsaap-Notes sur le processus d'évaluation d'audience nominale	89
18	L'interface de Tsaap-Notes dans un smartphone	90
19	Exemples de <i>scopes</i>	91
20	Annotation d'une diapositive avec Tsaap-Notes	92
21	Interface d'édition d'une question	93
22	Interface de soumission d'une réponse	94
23	Présentation des résultats	94
24	La fonction d'exportation des questions relatives à un <i>scope</i>	97
25	Le fichier généré par l'export des questions	98
26	Édition d'une question issue de Tsaap-Notes et importée dans Moodle	99

27	Affichage du feedback dans Moodle pour une réponse donnée par un étudiant	99
28	Soumission d'une réponse : première phase	101
29	Soumission d'une réponse : deuxième phase	102
30	Soumission d'une réponse : troisième phase	103
31	L'algorithme pour générer le conflit socio-cognitif	106
32	L'interface de Moodle présentant le feedback provenant de Tsaap . . .	107
33	Taux de participation sur les soumissions de réponses et sur les soumissions d'explications	125
34	Nombre de mauvaises réponses en phase 1 et en phase 2	126
35	Taux de participation à la tâche d'évaluation par les pairs	127

Liste des tableaux

1	Processus d'évaluation formative (Black et Wiliam, 2009)	25
2	Matrice Activité/Stratégie	26
3	Grille de critères d'évaluation des dispositifs étudiés	31
3	Grille de critères d'évaluation des dispositifs étudiés	32
3	Grille de critères d'évaluation des dispositifs étudiés	33
4	Synthèse de l'évaluation des systèmes de votes interactifs	44
5	Synthèse de l'évaluation des systèmes d'échange et de partage de textes	48
6	Synthèse de l'évaluation des systèmes étudiés	49
7	Éléments de base de BPMN 2.0 (OMG, 2011)	64
7	Éléments de base de BPMN 2.0 (OMG, 2011)	65
8	Évaluation de la plateforme Tsaap-Notes	109
9	Éléments clés de l'environnement d'exploitation	114
10	Statistiques sur l'activité de prise de notes	119
11	Impact des votes interactifs	119
12	Résultats obtenus sur les questions de type "Oui/Non"	120
13	Taux de participation aux réponses soumises en N phases	125
14	Les items du <i>System usability Scale</i>	128
15	Évaluation de la plateforme Tsaap-Notes	131

Introduction

Chapitre 1

Introduction

Sommaire

1.1	Cadre scientifique	3
1.2	Contexte	5
1.2.1	Contexte sociétal - une crise de l'enseignement supérieur ?	5
1.2.2	Contexte épistémologique	6
1.3	Problématique de recherche et contributions	8
1.3.1	Les questions de recherche	8
1.3.2	Méthodologie	9
1.3.3	Contributions	10
1.4	Organisation du mémoire	11
1.4.1	Partie I : "État de l'art"	11
1.4.2	Partie II : "Systèmes d'évaluation d'audience"	11
1.4.3	Partie III : "Expérimentations"	11

1.1 Cadre scientifique

Ce manuscrit présente la synthèse de trois ans de travail effectué dans le cadre d'une thèse. Celle-ci s'est déroulée au sein de l'équipe SIERA¹ du laboratoire IRIT² à Toulouse. L'IRIT est une Unité Mixte de Recherche (UMR 5505) commune à

1. Service IntEgration and netwoRk Administration (Administration de Réseaux Et Intégration de Services)

2. Institut de Recherche en Informatique de Toulouse

plusieurs établissements qui sont l'université Toulouse III Paul Sabatier³, le CNRS⁴, l'INPT⁵, et l'université Toulouse I Sciences Sociales⁶. Les priorités du laboratoire sont le développement (1) de projets transversaux ouverts aux communautés nationales et internationales, (2) d'actions de valorisation et de diffusion de la culture scientifique et technique, (3) de transferts de connaissances par l'organisation régulière de rencontres Recherche/Monde socio-économique et d'opérations en direction du grand public et des scolaires et enfin, (4) de transferts de savoir-faire et de technologies par la création de laboratoires communs avec l'industrie. Les travaux de l'équipe SIERA, dirigée par Abdelmalek Benzekri, visent le contrôle et la maîtrise des infrastructures et services de communication de dernières générations, mais aussi des systèmes et applications complexes dynamiquement agrégés, distribués et trans-organisationnels. Ainsi, ces travaux s'attachent à la définition et à l'évaluation de nouveaux paradigmes de gestion et apportent à la communauté des architectures, des plates-formes, des outils ou encore des contributions à la normalisation. Dans ce contexte, plusieurs thématiques particulières font l'objet d'études et de propositions : (1) une démarche de conception d'applications de gestion dégageant une double indépendance, l'une vis-à-vis des domaines spécifiques à gérer et l'autre vis-à-vis des plates-formes de gestion, (2) l'évaluation du potentiel des SMA⁷ à répondre aux exigences de la gestion coopérative et de la gestion autonome, (3) la gestion de la sécurité des flux d'information d'une organisation virtuelle à travers le raffinement de politiques pour concevoir et déployer des politiques inter-domaines, et la définition d'un environnement de gestion qui permet de contrôler les politiques de sécurité et de valider formellement les règles exécutables de sécurité implantées liées aux mécanismes et protocoles de sécurité appropriés, et (4) la distribution et la gestion des environnements de formation en ligne qui constituent de nos jours des systèmes complexes d'exploitation d'objets pédagogiques à grande échelle. C'est dans ce dernier thème que s'inscrit le sujet des travaux exposés dans ce manuscrit.

3. <http://www.univ-tlse3.fr>

4. Centre National de la Recherche Scientifique - <http://www.cnrs.fr>

5. Institut National Polytechnique de Toulouse - <http://www.inpt-toulouse.fr>

6. <http://www.univ-tlse1.fr>

7. Service Monitoring Agent

1.2 Contexte

1.2.1 Contexte sociétal - une crise de l'enseignement supérieur ?

En 2009, l'Institut de Statistiques de l'UNESCO pointe dans son recueil de données mondiales sur l'éducation (UNESCO, 2009) une croissance très importante du nombre d'étudiants dans le supérieur :

"Le nombre d'étudiants dans le supérieur a connu une croissance exceptionnelle au cours des 37 dernières années : il a en effet été multiplié par cinq durant cette période, passant de 28,6 millions en 1970 à 152,5 millions en 2007. Ceci se traduit par une croissance annuelle moyenne de 4,6%, tandis que le nombre moyen d'étudiants double tous les 15 ans. En examinant ces chiffres de plus près, ces derniers révèlent que cette croissance a été particulièrement prononcée depuis 2000, avec 51,7 millions de nouveaux inscrits dans l'enseignement supérieur dans le monde en 7 ans seulement."

Cette croissance du nombre d'étudiants dans le supérieur pose d'abord la question des moyens nécessaires à leur formation. L'UNESCO en vient à se demander si les gouvernements sont réellement en capacité de soutenir une telle croissance ou si celle-ci doit s'appuyer sur des fonds privés (UNESCO, 2009). Dans la plupart des pays les étudiants qui remplissent les amphithéâtres ont des profils différents de ceux qui peuplaient les universités il y a 40 ans. En France par exemple, la politique lancée en 1985 ayant mené 80% d'une même classe d'âge au niveau du baccalauréat a attiré vers l'université des profils qui n'avaient pas accès à l'enseignement supérieur quelques années auparavant. Enfin, l'exposition aux nouvelles technologies de toutes les générations nées après 1981 (génération Y) a altéré de manière significative la manière de vivre des nouveaux étudiants et plus particulièrement leur manière d'appréhender le travail (Aksoy *et al.*, 2013).

La croissance importante du nombre d'étudiants et le changement des profils d'apprenants ont eu un impact important sur les résultats obtenus dans l'enseignement supérieur. En France, par exemple, les chiffres catastrophiques concernant le taux d'échec en licence a mené au lancement en 2007 du Plan réussite en licence (PLR) ayant pour objectif de mener en 5 ans 50% d'une classe d'âge au niveau licence. Pour atteindre cet objectif, le Ministère de l'Éducation Nationale,

de l'Enseignement Supérieur et de la Recherche (MENESR) fournit une allocation pluriannuelle de crédits supplémentaires pour aider les universités à développer leur action en faveur de la réussite des étudiants engagés dans un cursus de licence. En 2014, le Plan réussite en licence n'a pas atteint ses objectifs comme l'indique le MENESR dans son rapport annuel 2014 "Repères et références statistiques" (MENESR, 2014) :

"27,3% des étudiants inscrits pour la première fois en L1 en 2008-2009 ont obtenu leur diplôme trois ans plus tard à la session 2011. Une année supplémentaire a été nécessaire pour 12,4% des inscrits. Au total, 39,7% des étudiants inscrits en 2008 ont obtenu leur licence en trois ou quatre ans. Les étudiants inscrits en 2009-2010 ont un taux de réussite en trois ans moins élevé que celui de la cohorte 2008 (26,3%). Entre la cohorte 2004 et la cohorte 2009, le taux de réussite en trois ans a perdu 2,8 points, notamment sous l'effet de la moindre réussite des titulaires d'un baccalauréat général sur la période (-3,8 points pour les bacheliers littéraires, -4,2 pour la série ES et -1,4 point pour les scientifiques)."

De manière concrète, certaines universités comme celle de Toulouse III ont utilisé une partie des moyens du PLR pour remplacer les cours de licence en amphithéâtre pouvant accueillir plusieurs centaines d'étudiants par des cours/TD accueillant au maximum une quarantaine d'étudiants, espérant obtenir ainsi de meilleurs résultats. L'impact sur les coûts fut considérable malgré l'allocation du MENESR et les résultats n'ont pas été probants. En conséquence l'université de Toulouse III a décidé pour la prochaine accréditation un retour aux cours en amphithéâtres en promouvant les pédagogies mettant en situation d'activité les étudiants (pédagogies dites actives). L'utilisation de boitiers de votes interactifs pendant les cours en amphithéâtres est plébiscitée.

1.2.2 Contexte épistémologique

Bloom, dans son article de 1984 intitulé "The 2 Sigma Problem : the search for methods of group instruction as effective as one-to-one tutoring" (Bloom, 1984), rend compte d'une expérience ayant pour objet la comparaison de résultats d'apprentissage obtenus par des apprenants dans le contexte d'un enseignement en classe (un enseignant pour un groupe d'une trentaine d'apprenants) avec ceux obtenus par des apprenants dans le contexte d'un enseignement avec un tuteur

(un enseignant pour un apprenant). Les résultats de cette expérience synthétisés dans la Figure 1 montrent que l'apprentissage en situation de tutorat ("un à un") conduit à des résultats très remarquablement supérieurs à ceux obtenus en situation d'enseignement en classe. Considérant que le coût de mise en œuvre d'une politique de généralisation de l'enseignement "un à un" n'est pas soutenable à une très grande échelle (celle d'un pays par exemple), Bloom appelle à un effort de recherche pour trouver des solutions permettant d'obtenir d'un enseignement en classe des résultats similaires à ceux obtenus en situation de tutorat "un à un". Cet appel de 1984 reste criant d'actualité dans le contexte de l'enseignement supérieur décrit dans le paragraphe précédent.

Suite à l'appel de Bloom, la recherche en informatique s'est mobilisée et structurée autour de quelques programmes de recherche au sens de Lakatos (Lakatos *et al.*, 1994).

S'appuyant sur les théories de l'Intelligence artificielle côté informatique et sur les théories de l'apprentissage plutôt cognitivistes, les programmes de recherche portant sur les Systèmes Tutoriels Intelligents (STI) visent le développement de logiciels pouvant se substituer à un tuteur humain. La recherche sur les STI produit des résultats mais dans un cadre très contraint : (1) chaque STI mis en œuvre ne couvre qu'un domaine de connaissances ciblé (tel que le SQL avec le SQL Tutor (Mitrovic, 2010), la géométrie avec le Geometry Tutor (Anderson *et al.*, 1985), etc.), (2) les STI nécessitent un environnement technique dédié (un STI développé avec le logiciel

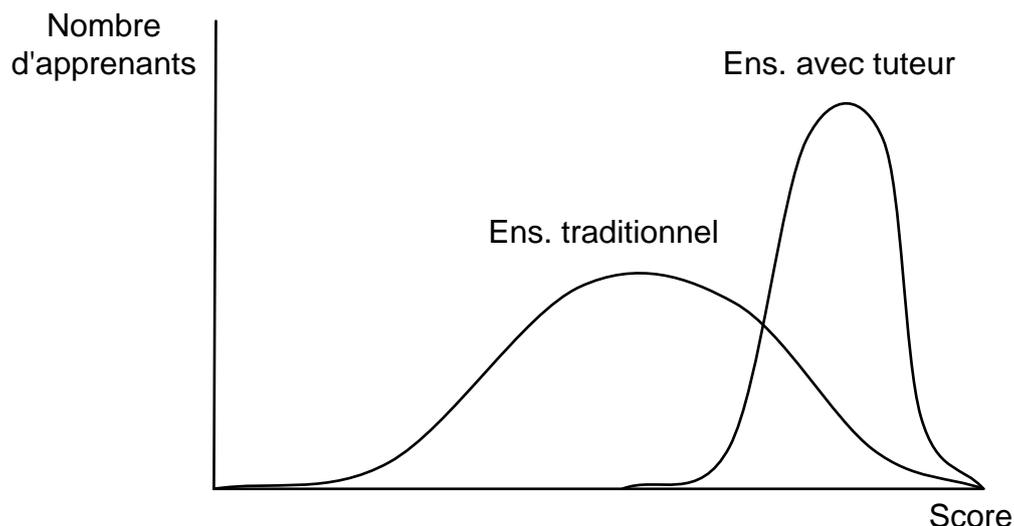


Figure 1. Synthèse des résultats obtenus par Bloom

CTAT nécessite l'environnement CTAT pour fonctionner (Aleven *et al.*, 2006)), et (3) les STI sont peu ouverts sur les environnements technologiques existants tels que les Learning Management Systems (LMS) ou les Environnements Numériques de Travail (ENT) pourtant massivement utilisés dans les établissements de formation. D'autre part, la complexité et le coût de mise en œuvre de ces solutions sont encore trop importants (Aleven, 2010). Il n'existe donc à ce jour des STI que pour très peu de disciplines et très peu de niveaux, et le temps long de la recherche sur les STI n'est pas compatible avec les changements rapides qui bouleversent l'enseignement supérieur.

Nous proposons de travailler sur un programme de recherche se fondant davantage sur des théories de l'apprentissage socio-constructivistes en réconciliant ainsi l'apprenant avec le groupe. Cet angle de recherche, associé à une démarche de type "recherche action" caractérisée entre autres par l'objectif dual de résoudre le problème des usagers et de faire avancer des connaissances fondamentales (Liu, 1997), positionnent le contexte épistémologique de nos travaux.

1.3 Problématique de recherche et contributions

1.3.1 Les questions de recherche

L'enseignement dispensé à un groupe composé d'une quarantaine d'individus ou plus est toujours d'actualité dans l'enseignement supérieur car il est un des dispositifs les moins coûteux et donc un des dispositifs les plus soutenables à très grande échelle. La question de l'amélioration et de l'adaptation de la pratique d'enseignement dans un tel cadre pour le nouveau public présent dans l'enseignement supérieur est au cœur de notre recherche.

Les systèmes de votes interactifs (SVI) font l'objet d'un grand nombre d'études (Caldwell, 2007; Kay et LeSage, 2009) démontrant les bénéfices de leur utilisation. Les SVI maximisent l'engagement des étudiants (Gauci *et al.*, 2009; Uhari *et al.*, 2003), et les résultats présentés sous forme d'histogrammes donnent à l'enseignant l'opportunité de déclencher et d'arbitrer des discussions constructives entre les étudiants (Gauci *et al.*, 2009). En synthèse, les SVI apparaissent comme un moyen d'introduire de l'évaluation formative dans les amphithéâtres (Kay et LeSage, 2009). Les modalités d'évaluations formatives proposées par les SVI sont cependant limitées par le faible niveau d'interaction autorisé par ces dispositifs : les questions posées

à l'aide des SVI ne tolèrent pas de réponses saisies sous forme de texte libre, et les productions écrites ne peuvent donc pas être évaluées dans les processus induits par l'utilisation des SVI. Les contraintes sur les types d'interaction impactent directement la nature des feedbacks fournis par les SVI qui se cantonnent à des indicateurs numériques.

L'apparition dans les amphithéâtres de dispositifs équipés de claviers et connectés à Internet a rendu de nouvelles interactions possibles : les étudiants peuvent utiliser leur smartphone, leur tablette ou ordinateur portable pour échanger ou partager des documents ou messages textuels. Ces nouvelles possibilités ont ouvert la porte à de nouveaux usages dans les amphithéâtres tels que la prise de notes collaborative (Kam *et al.*, 2005) ou le microblogging (Ebner, 2009). Ces usages conduisent à des bénéfices importants pour les étudiants qui peuvent ainsi recevoir des feedbacks produits par les pairs exprimés sous forme textuelle.

Il semble donc possible de tirer partie des fonctionnalités offertes par ces systèmes d'échange et de partage de textes pour dépasser les limites des SVI actuels et concevoir un système d'évaluation formative enrichi utilisable dans les amphithéâtres d'universités. Nos questions de recherche peuvent donc être formulées ainsi :

1. Comment enrichir la mise en œuvre d'évaluations formatives dans le contexte d'un cours en face à face dispensé à un grand nombre d'étudiants ?
2. Comment susciter l'engagement du plus grand nombre d'étudiants dans les activités relatives à ces évaluations formatives ?

1.3.2 Méthodologie

Afin d'adresser les questions de recherche sus-citées, nous avons conçu et adopté la méthode composée des phases suivantes :

1. l'analyse théorique de la notion d'évaluation formative, ainsi que l'analyse des systèmes existants visant la mise en œuvre de ce type d'évaluations afin d'explicitier les verrous à lever ;
2. la définition d'une stratégie et d'un nouveau processus pour combler les lacunes identifiées ;
3. la mise en œuvre d'un système implantant la stratégie et le nouveau processus ;
4. la mise en place d'expérimentations dans un contexte réel.

1.3.3 Contributions

Nos contributions portent sur quatre différents aspects, à la fois théoriques et technologiques.

1.3.3.1 La définition du concept d'évaluation d'audience

L'introduction de ce nouveau concept et la description des principaux processus qui le caractérisent permettent de définir de manière claire et précise le périmètre de nos recherches, et offre un cadre plus large pour de recherches futures. Notre système cible enrichissant la mise en œuvre de l'évaluation formative dans le contexte des cours en face à face peut être qualifié de "système d'évaluation d'audience".

1.3.3.2 La modélisation standardisée des processus d'évaluation d'audience

Afin de faciliter la lecture, la compréhension et les échanges au sein de la communauté des acteurs concernés par l'évaluation d'audience, nous avons modélisé les processus les plus représentatifs ainsi que le processus en N phases à l'aide du langage standard BPMN 2.0 spécifié par l'Object Management Group⁸.

1.3.3.3 La spécification d'un nouveau processus d'évaluation d'audience

L'enrichissement de la mise en œuvre d'évaluations d'audience a nécessité la conception d'un nouveau processus baptisé "processus en N phases". Celui-ci, en introduisant le support de nouvelles activités et stratégies caractéristiques de l'évaluation formative, représente la clé de voûte des moyens mis en œuvre pour lever les verrous identifiés.

1.3.3.4 La conception et la mise en œuvre d'un système d'évaluation d'audience

Le système Tsaap-Notes conçu et développé au cours de nos travaux est un système d'évaluation d'audience implantant le processus en N phases. Tsaap-Notes enrichit de fait la mise en œuvre de l'évaluation formative par rapport aux systèmes existants. Il parvient également à engager la très grande majorité des étudiants qui assistent aux cours utilisant le système.

8. <http://www.omg.org/spec/BPMN/2.0/>

1.4 Organisation du mémoire

La suite du document est composée de 3 parties principales ; les conclusions et perspectives de recherche clôturent le manuscrit.

1.4.1 Partie I : "État de l'art"

Cette partie est composée de deux chapitres. Le premier chapitre est une revue de la littérature scientifique portant sur l'évaluation formative. Son objectif est de fournir le cadre théorique dans lequel nous avons conçu la grille d'évaluation des dispositifs technologiques étudiés dans le chapitre suivant.

Le deuxième chapitre présente l'étude des deux familles de systèmes utilisés aujourd'hui dans les amphithéâtres : les systèmes de votes interactifs et les systèmes d'échange et de partage de textes. Ces systèmes sont étudiés et évalués à l'aide de la grille élaborée précédemment. Ce chapitre s'achève par la spécification des verrous que nous proposons de lever dans le cadre de nos travaux de recherche.

1.4.2 Partie II : "Systèmes d'évaluation d'audience"

Cette partie détaille les différentes contributions produites au cours de nos travaux de recherche. Elle est composée de deux chapitres. Le premier chapitre introduit le concept d'évaluation d'audience et les processus les plus représentatifs associés au concept.

Le deuxième chapitre présente la conception et la mise en œuvre d'un système d'évaluation d'audience : la plateforme Tsaap-Notes.

1.4.3 Partie III : "Expérimentations"

Cette partie présente le contexte et les résultats obtenus lors des trois expérimentations menées durant trois années universitaires sur différents stades d'évolution du système exposé dans la partie précédente.

Nous terminons le mémoire par un chapitre de conclusion présentant la synthèse des contributions apportées par notre travail et ouvrant sur différentes perspectives de recherche.

Première partie

État de l'art

Chapitre 2

L'évaluation formative

Sommaire

2.1	Introduction	15
2.2	Évaluation formative vs Évaluation sommative	16
2.3	Le feedback	17
2.3.1	Définition	17
2.3.2	Indicateurs numériques vs commentaires textuels	18
2.3.3	Les conditions d'un feedback efficace	18
2.4	L'auto-régulation	19
2.5	Auto-évaluation et évaluation par les pairs	21
2.5.1	Auto-évaluation	21
2.5.2	Évaluation par les pairs	22
2.6	Synthèse : éléments de la théorie de l'évaluation formative	23
2.6.1	Nouvelle définition de l'évaluation formative	23
2.6.2	Activités de référence pour l'évaluation formative	24
2.6.3	Stratégies caractéristiques de l'évaluation formative	24
2.7	Conclusion	26

2.1 Introduction

L'étude exhaustive publiée en 1998 par Black et Wiliam (1998), confortée par d'autres études (Bloom, 1984; Kluger et DeNisi, 1996; Meisels, 2003; Rodriguez,

2004), a démontré que l'évaluation formative est un dispositif d'apprentissage efficace et ceci quels que soient le champ disciplinaire, le type de connaissances et de compétences visées, et quel que soit le niveau de formation visé. Cette étude met définitivement l'évaluation formative au premier rang des pratiques d'apprentissage à solliciter. Il émerge de cette étude le caractère essentiel du feedback¹ dans les situations observées ainsi que l'importance des capacités d'auto-régulation des étudiants. Ce chapitre présente un panorama des concepts clés de l'évaluation formative afin d'en tirer les éléments fournissant le cadre théorique de nos travaux de recherche.

2.2 Évaluation formative vs Évaluation sommative

Sadler (1989) définit l'évaluation formative comme évaluation qui s'intéresse à l'utilisation des évaluations des réponses des étudiants (performances, essais, devoirs) pour façonner et améliorer les compétences des étudiants en court-circuitant le caractère hasardeux et inefficace d'un apprentissage reposant uniquement sur une approche "essai-erreur".

De manière plus synthétique, dans son projet pour une approche pragmatique de l'évaluation formative (Perrenoud, 1989), Perrenoud définit l'évaluation formative ainsi :

"Est formative toute évaluation qui aide l'élève à apprendre et à se développer, autrement dit, qui participe à la régulation des apprentissages et du développement dans le sens d'un projet éducatif."

Les évaluations formatives se distinguent des évaluations sommatives ayant pour objectif la mesure du niveau d'atteinte par les étudiants des objectifs d'un programme scolaire (Bloom, 1971). Les évaluations sommatives permettent, à la fin d'une unité d'instruction, de comparer le niveau de l'étudiant à un niveau de référence attendu (Eberly Center, 2015) en vue de l'éventuelle délivrance d'une certification (Sadler, 1989). Wiliam (2007) introduit les expressions "évaluation pour l'apprentissage" et "évaluation de l'apprentissage"² pour distinguer respectivement

1. Suivant les traces de Perrenoud, nous conserverons le terme anglo-saxon dans la suite du document.

2. "assessment for learning" et "assessment of learning" dans le texte.

évaluation formative et évaluation sommative.

Plus récemment, Eberly Center (2015) définit l'évaluation formative de la manière suivante :

"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 adresser les problèmes immédiatement."

Cette définition explicite le rôle clé joué par le concept de feedback au cœur de l'évaluation formative, et corrobore l'idée que les concepts de feedback et d'évaluation formative se chevauchent fortement (Black et Wiliam, 1998).

2.3 Le feedback

2.3.1 Définition

Sadler (1989) lie le terme feedback employé dans le contexte de l'évaluation formative à la définition donnée par Ramaprasad (1983) dans le contexte de la théorie du management :

"Un feedback est une information portant sur l'écart entre un niveau actuel et le niveau de référence d'un paramètre d'un système qui est utilisée pour modifier cet écart d'une quelconque manière."

Dans le contexte de l'évaluation formative, le paramètre du système est le niveau d'atteinte par un étudiant d'un ou de plusieurs objectifs d'une unité d'instruction.

Alors que pour Kluger et DeNisi (1996) la seule présence de l'information sur l'écart entre le niveau actuel et le niveau de référence constitue le feedback, Ramaprasad et Sadler considèrent tous deux que l'information sur l'écart n'est un feedback que lorsque cette information est utilisée pour modifier l'écart. En d'autre terme, c'est l'usage qui est fait d'une information qui la caractérise comme feedback. Sous l'angle de la théorie de la genèse instrumentale (Rabardel, 1995), l'information

est un *artefact* qui ne devient *instrument* pour aider l'étudiant que si l'étudiant et/ou l'enseignant en font un usage explicite dans ce sens.

Dans la suite du document, nous utiliserons le terme feedback pour désigner aussi bien l'artefact que l'instrument.

Hattie et Timperley (2007) définissent le feedback comme une information fournie par un agent (e.g. un enseignant, un pair, un parent, l'apprenant lui-même, une expérience) concernant les aspects d'une performance ou d'une compréhension d'un apprenant. De manière très synthétique, les auteurs perçoivent le feedback comme une "conséquence" à une performance.

2.3.2 Indicateurs numériques vs commentaires textuels

Plusieurs études (Black et Wiliam, 1998; Crooks, 1988; Butler, 1987) démontrent l'efficacité supérieure d'un feedback exprimé sous la forme de commentaires textuels par rapport à un feedback exprimé sous la forme d'un indicateur numérique (note chiffrée, score, pourcentage, échelle de Lickert). Butler (1987) indique que les scores peuvent augmenter l'engagement des étudiants mais qu'ils n'améliorent pas leurs performances. Dans cette même étude, Butler montre que les feedbacks fournis sous forme textuelle uniquement ont conduit à des gains de performance, contrairement aux feedbacks fournis sous forme de scores uniquement ou sous forme de scores avec commentaires textuels.

Cette distinction dans la nature des feedbacks exprimés apparaît donc comme un critère discriminant dans les manières de mettre en œuvre l'évaluation formative.

2.3.3 Les conditions d'un feedback efficace

Sadler (1989) identifie trois conditions nécessaires pour que les étudiants tirent un bénéfice d'un feedback dans le contexte d'une tâche d'apprentissage ; il déclare qu'un étudiant doit connaître :

1. ce qu'est une bonne performance (i.e. il doit posséder le concept de l'objectif ou de la norme à atteindre) ;
2. comment se positionne sa performance actuelle par rapport à la bonne performance ;
3. comment agir pour réduire l'écart entre sa performance actuelle et la bonne performance.

Sadler pointe alors le fait que les étudiants doivent être capables d'interpréter le feedback et de faire la connection entre le feedback reçu et les caractéristiques du travail qu'ils ont produit. Il indique que la condition indispensable pour qu'il y ait progrès est que les étudiants arrivent à apprécier la qualité de leur travail à peu près comme un enseignant est capable de le faire - capacité d'auto-évaluation - et qu'ils soient capables de réguler en continu la qualité de ce qu'ils produisent au moment même où ils produisent. Sadler définit la combinaison de ces capacités comme la capacité d'auto-régulation. L'auto-régulation apparaît donc comme partie intégrante de l'évaluation formative.

2.4 L'auto-régulation

Sadler (1989) met en lumière les liens étroits entre feedback et auto-régulation : dans le cadre d'une évaluation formative, quand l'information provient d'un agent extérieur, il s'agit d'un feedback ; quand il s'agit d'une information générée par l'étudiant, il s'agit d'une auto-régulation. Les études postérieures à Sadler (1989) désignent cette information générée par l'étudiant par l'expression *feedback interne* et définissent de manière plus formelle le processus d'auto-régulation (Butler et Winne, 1995; Nicol et Macfarlane-Dick, 2006).

La Figure 2 illustre le modèle d'apprentissage auto-régulé proposé par Nicol et Macfarlane-Dick (2006). Le système cognitif de l'apprenant dans le contexte d'un apprentissage auto-régulé est composé d'un ensemble de connaissances et de croyances. Ce lot de connaissances et de croyances, couplé à l'activité d'auto-régulation, permet à l'étudiant d'établir des objectifs relatifs à une tâche à accomplir, de définir les stratégies et tactiques pour atteindre ces objectifs, et finalement de produire un résultat observable par l'extérieur. Ce résultat fait l'objet d'une mesure de performance externe menant à la fourniture d'un feedback externe relatif à la tâche à accomplir. Ce feedback, pouvant être fourni par l'enseignant, par un pair ou par tout autre dispositif (comme un ordinateur), modifie le lot de connaissances et de croyances de l'apprenant conduisant à un nouveau cycle de définition d'objectifs, de stratégies et de productions jusqu'à l'atteinte de la performance attendue. Dans ce modèle, le feedback interne intervient à différents niveaux. Il est généré d'une part en réaction au feedback externe, condition indispensable pour que le feedback externe ait une influence sur l'apprentissage (Ivanic *et al.*, 2000) et d'autre part durant le processus d'auto-régulation pendant

lequel l'étudiant compare l'avancement de son travail (résultats internes) au regard des objectifs qu'il s'est fixé.

Nicol et Macfarlane-Dick (2006), s'inspirant en partie des conditions énoncées dans Sadler (1989), proposent dans leur étude sept principes incarnant des bonnes pratiques permettant de faciliter l'auto-régulation :

1. clarifier ce qu'est une bonne performance ;
2. faciliter le développement de l'auto-évaluation (réflexion) dans l'apprentissage ;

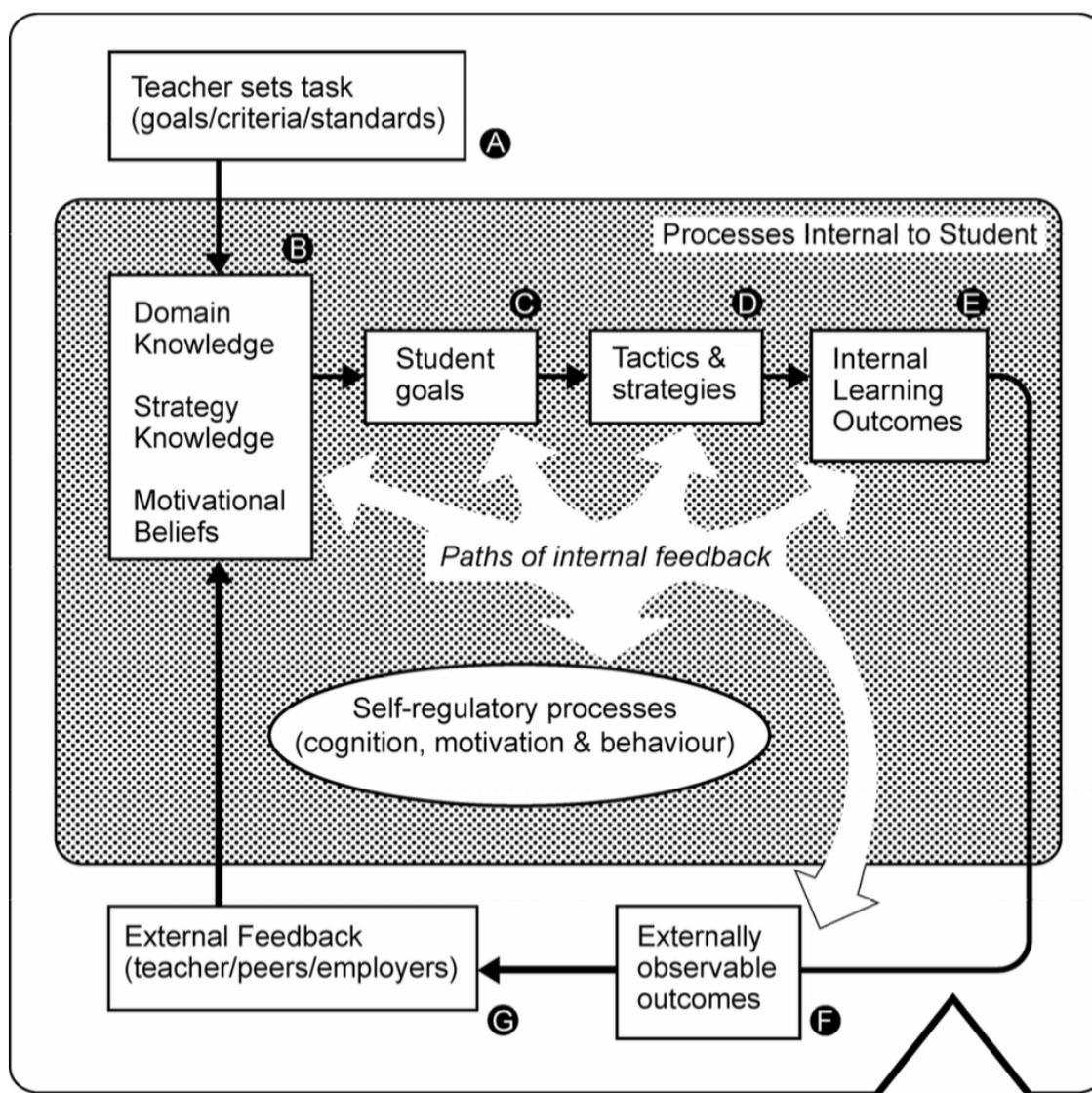


Figure 2. Modèle d'apprentissage auto-régulé (Nicol et Macfarlane-Dick, 2006)

3. fournir des informations de haute qualité aux étudiants à propos de leur apprentissage ;
4. encourager le dialogue entre enseignants et étudiants, entre pairs à propos de l'apprentissage ;
5. encourager les croyances motivationnelles positives et l'estime de soi ;
6. fournir les opportunités de combler l'écart entre la performance actuelle et la performance de référence ;
7. fournir l'information permettant à l'enseignant de façonner son enseignement.

Le développement de la capacité des étudiants à s'auto-évaluer réapparaît comme principe clé. Pour s'auto-évaluer, les étudiants doivent être capables d'évaluer une production comme le ferait leur enseignant (Sadler, 1989). Proposer aux étudiants des expériences les conduisant à développer leur capacité à évaluer la qualité d'un travail devient alors un pré-requis à toute volonté de développer les capacités d'auto-régulation.

2.5 Auto-évaluation et évaluation par les pairs

L'engagement des étudiants dans la pratique de l'évaluation peut prendre la forme de l'auto-évaluation, de l'évaluation par les pairs et de la co-évaluation (Falchikov, 1995). La co-évaluation laissant à l'enseignant l'autorité sur l'évaluation finale est en général pratiquée dans le contexte des évaluations sommatives (Dochy *et al.*, 1999), alors que l'auto-évaluation et l'évaluation par les pairs sont en général sollicitées dans le contexte de l'évaluation formative.

2.5.1 Auto-évaluation

Dans leur revue de 1999, Dochy *et al.* (1999) définissent l'auto-évaluation comme l'engagement des apprenants à émettre des jugements sur leur propre apprentissage, particulièrement sur leurs productions et sur les résultats de leur apprentissage. L'auto-évaluation est un moyen d'augmenter la participation des étudiants dans leur propre apprentissage (Boud, 2013). Dochy *et al.* confirment que les auto-évaluations sont principalement utilisées dans le contexte des évaluations formatives afin de favoriser la réflexion des étudiants sur leur propre processus et résultats d'apprentissage.

La revue (Dochy *et al.*, 1999) indique que les étudiants sont globalement capables de fournir des évaluations sur leurs productions très similaires aux évaluations proposées par les enseignants. Cependant l'étude pointe l'influence de certains caractères individuels : les meilleurs étudiants ont tendance à sous-évaluer leur production alors que les étudiants plus faibles ont tendance à sur-évaluer leur production. La qualité des prédictions de performances augmente avec le niveau de formation.

2.5.2 Évaluation par les pairs

Dans la revue Topping (1998), l'évaluation par les pairs est définie comme le procédé dans lequel des individus évaluent la quantité, le niveau, la valeur, la qualité ou le succès des productions ou des résultats d'apprentissages de pairs de statuts similaires. Cette définition permet d'exclure les pratiques consistant à rémunérer des étudiants d'un niveau supérieur pour l'évaluation d'étudiants d'un niveau inférieur.

L'étude de 31 publications sur l'évaluation par les pairs (Topping, 1998) indique que pour 18 études, l'évaluation par les pairs s'avère suffisamment fiable et valide ; seulement 7 études ont conclu à un niveau de fiabilité et de validité très faible. Dans tous les cas, Devenney (1989) pointe le fait que la fonction d'évaluation formative portée par l'évaluation par les pairs ne nécessite pas forcément un haut niveau de fiabilité.

En terme d'efficacité, la revue de la littérature de Topping (1998) permet de conclure sur le fait que la simple utilisation de feedbacks quantitatifs (notes chiffrées données à l'aide d'échelles de Lickert) suffit à obtenir des effets positifs sur l'apprentissage tant en terme d'amélioration des résultats aux tests sommatifs qu'en terme de perception subjective des participants. Topping (1998) relève néanmoins le manque d'étude prenant en compte les feedback qualitatifs (commentaires textuels) émis par les pairs.

Notons que ce dernier constat exprimé par Topping peut paraître en contradiction avec les résultats présentés dans le paragraphe 2.3.2. Les gains d'apprentissage présentés par Topping sont observés dans le contexte particulier de l'évaluation par les pairs où les apprenants sont mis en situation d'évaluateurs. Il semble donc que cette position d'évaluateur, plus que le type de feedback émis, permet de rendre compte des gains observés.

2.6 Synthèse : éléments de la théorie de l'évaluation formative

Dans la continuité de leur travaux sur plus de deux décennies portant sur l'évaluation formative, Black et Wiliam (2009) proposent une base théorique unifiée de l'évaluation formative. Ce travail conduit à la proposition d'un cadre théorique de l'évaluation formative s'appuyant sur une nouvelle définition.

2.6.1 Nouvelle définition de l'évaluation formative

La définition proposée par Black et Wiliam (2009) met en avant divers aspects de l'évaluation formative déjà présents dans les précédentes définitions mais apportent de nouvelles précisions. Ces nouveaux éléments s'avèrent précieux d'une part pour la constitution d'un cadre théorique de l'évaluation formative, et d'autre part pour sa mise en œuvre concrète.

"Une pratique dans une classe relève de l'évaluation formative quand des preuves concernant le niveau de succès³ des étudiants sont élicitées, interprétées et utilisées par les enseignants, les apprenants, ou leurs pairs, pour prendre des décisions sur les prochaines étapes d'instruction qui seront probablement meilleures, ou davantage fondées que des décisions qui auraient été prises en l'absence des preuves élicitées."

Comme cela était déjà sous-tendu dans les précédentes définitions de l'évaluation formative, les preuves collectées doivent être utilisées en terme de prises de décisions concrètes pour qu'il y ait évaluation formative. Le fait que la prise de décision ne soit pas l'apanage de l'enseignant est clairement explicité en mettant en avant le rôle de l'apprenant mais aussi de ses pairs dans ce champ de responsabilité. Cette définition introduit aussi la notion d'incertitude sur les résultats obtenus en utilisant l'adverbe "probablement" comme complément de l'adjectif "meilleur". Les auteurs insistent sur ce parti pris en rappelant que les interventions les mieux conçues ne mènent pas toujours tous les étudiants à un meilleur apprentissage. Enfin, l'idée que les preuves collectées permettent de prendre des décisions "probablement meilleures" ou "davantage fondées" prend en compte les cas où les informations collectées remettent en question une stratégie d'enseignement ou d'apprentissage, mais aussi les cas où ces informations confortent la stratégie initiale.

3. le terme *achievement* est utilisé par les auteurs dans le texte original

Dans la suite du document, l'expression "évaluation formative" fera référence à cette nouvelle définition proposée par Black et Wiliam (2009).

2.6.2 Activités de référence pour l'évaluation formative

Le point de départ du modèle de Black et Wiliam (2009) est la synthèse de travaux antérieurs (Wiliam, 2000, 2007; Black *et al.*, 2003) menant à l'identification de cinq activités pratiquées dans les classes pour la mise en œuvre de l'évaluation formative :

- le partage des critères de succès avec les apprenants (A1) ;
- le questionnement à la classe (A2) ;
- l'évaluation sous forme de commentaires plutôt qu'à l'aide de notes chiffrées (A3) ;
- l'auto-évaluation et évaluation par les pairs (A4) ;
- l'utilisation de tests sommatifs en vue d'usage formatif (A5).

Dans la suite du document, nous désignerons ces cinq activités comme les "activités de référence" de l'évaluation formative.

Bien que ces activités de référence apparaissent liées à l'idée centrale d'évaluation formative, la manière précise dont elles sont liées n'est pas clairement définie ; de plus, il est difficile d'affirmer que ces activités suffisent à couvrir la totalité du champ de l'évaluation formative (Black et Wiliam, 2009). Le modèle théorique présenté dans le paragraphe suivant vise à dissiper ces imprécisions et incertitudes.

2.6.3 Stratégies caractéristiques de l'évaluation formative

Le cadre théorique proposé par Black et Wiliam (2009) s'inscrit dans les trois processus clés de l'apprentissage et de l'enseignement (Dwyer, 2008) généralisant les assertions de Ramaprasad (1983) sur le feedback :

- établir où se situent les apprenants dans leur apprentissage ;
- établir les objectifs d'apprentissage ;
- établir ce qui doit être fait pour que les apprenants atteignent les objectifs.

La déclinaison de ces trois processus suivant les différents acteurs (enseignant, apprenant, pairs) permet d'explicitier cinq stratégies qui caractérisent l'évaluation formative (Black et Wiliam, 2009) :

1. clarifier et partager les objectifs d'apprentissage et les critères de succès (S1) ;

Table 1. Processus d'évaluation formative (Black et Wiliam, 2009)

	Enseignant	Pairs	Apprenant
Quels sont les objectifs d'apprentissage ?	1. Clarifier les objectifs et les critères de succès	Comprendre et partager les objectifs et critères de succès	Comprendre les objectifs et critères de succès
Où se trouve l'apprenant ?	2. Concevoir des situations efficaces de discussions en classe et d'autres tâches d'apprentissage qui élicitent les preuves du niveau d'avancement des apprenants	4. Faire en sorte que les apprenants deviennent des ressources d'instruction pour les autres	5. Faire en sorte que les apprenants deviennent responsables de leur propre apprentissage
Quoi faire pour que l'apprenant atteigne les objectifs ?	3. Fournir des feedbacks qui permettent aux apprenants de progresser		

2. concevoir des situations efficaces de discussions en classe et d'autres tâches d'apprentissage qui élicitent les preuves du niveau d'avancement des apprenants dans leur apprentissage (S2) ;
3. fournir des feedbacks qui permettent aux apprenants de progresser dans leur apprentissage (S3) ;
4. faire en sorte que les apprenants deviennent des ressources d'instruction pour les autres apprenants (S4) ;
5. faire en sorte que les apprenants deviennent responsables de leur propre apprentissage (S5).

Le Tableau 1 présente les cinq stratégies et leur articulation avec les autres composantes des différents processus mentionnés ci-dessus.

Les activités de référence présentées dans la section précédente peuvent alors être utilisées pour implanter les différentes stratégies dans les processus se déroulant concrètement dans la classe. Le lien entre l'activité A1 et la stratégie S1 est trivial. Le questionnement de la classe (A2) est un moyen de proposer une situation de discussion en classe (S2) ; l'utilisation d'évaluations sous forme de commentaires (A3) est un moyen de communiquer du feedback (S3) ; auto-évaluation et évaluation

Table 2. Matrice Activité/Stratégie

	S1	S2	S3	S4	S5
A1					
A2					
A3					
A4					
A5					

par les pairs (A4) s'inscrivent dans les stratégies (S4) et (S5). Enfin, l'utilisation de tests sommatifs à des fins formatives (A5) est un moyen de clarifier des objectifs à atteindre et des critères de succès (S1) ; dans le contexte de l'élicitation du niveau de succès des étudiants (S2), les tests sommatifs utilisés comme tests formatifs représentent un levier pour la délivrance de feedback (S3). Les liens entre types d'activités et stratégies qui caractérisent l'évaluation formative sont synthétisés dans une matrice à deux dimensions représentée dans le Tableau 2, où une case ($A_i ; S_j$) grisée représente une activité A_i participant à la stratégie S_j .

2.7 Conclusion

Nos travaux de thèse s'inscrivent dans le champ disciplinaire des Environnements Informatiques pour l'Apprentissage Humain (EIAH). Ce premier chapitre d'état de l'art situe nos travaux dans sa dimension "apprentissage humain". L'évaluation formative est au cœur de notre recherche. Il reste maintenant à situer nos travaux sur la dimension "environnement informatique".

En 2003, Yorke (2003) constate la sous utilisation des évaluations formatives dans l'enseignement supérieur et identifie plusieurs causes à ce phénomène :

- une préoccupation croissante aux normes de rendements scolaires conduisant à mettre l'accent sur les évaluations (sommatives) de résultats ;
- une augmentation du ratio étudiants/enseignants menant à la diminution de l'attention apportée à chaque individu ;
- une augmentation et diversification des activités exigées aux enseignants du supérieur laissant de fait moins de temps à la mise en place d'évaluations formatives.

Reprenant les arguments de Shepard (2000), Yorke déclare que, bien que les méthodes d'apprentissage aient migré en direction des théories constructivistes,

les méthodes d'évaluation sont encore attachées aux activités de tests relevant davantage des théories béhavioristes.

À l'issue de ce chapitre, notre objectif principal peut être formulé de la manière suivante :

Concevoir et mettre en œuvre un système d'évaluation formative utilisable dans un amphithéâtre d'université supportant les stratégies et activités de références définies dans le cadre théorique de Black et Wiliam (2009).

Dès 1993, plusieurs travaux (Clariana, 1993; Zakrzewski et Bull, 1998) montrent l'efficacité de dispositifs d'auto-évaluation informatisés pour la mise en œuvre d'évaluations formatives. Les années 2000 voient apparaître un grand nombre de publications (Caldwell, 2007) démontrant les capacités des systèmes de votes interactifs à mettre en œuvre de l'évaluation formative dans l'enseignement secondaire ou supérieur. Le chapitre suivant étudie les apports de différentes technologies dans la sphère de l'évaluation formative pratiquée dans l'enseignement supérieur. Dans cette étude, nous analysons les dispositifs technologiques sous le spectre d'une grille de critères intégrant les stratégies et activités de référence décrites précédemment. Cette étude dresse alors un état de l'art des systèmes existants afin d'identifier d'une part les limites de ces systèmes au regard de la théorie de l'évaluation formative de Black et Wiliam (2009), et d'autre part les verrous à lever pour la conception et la mise en œuvre d'un système d'évaluation formative dépassant les limites des systèmes actuels.

Chapitre 3

Technologies et évaluation formative dans les amphithéâtres

Sommaire

3.1	Introduction	30
3.2	La méthode d'analyse	30
3.2.1	Les dispositifs retenus dans l'étude	30
3.2.2	La grille d'évaluation	31
3.2.3	Le processus d'analyse	36
3.3	Les Systèmes de votes interactifs	37
3.3.1	Principes généraux	37
3.3.2	Activités associées	38
3.3.3	Interactions et feedbacks	41
3.3.4	Engagement des étudiants	43
3.3.5	Évaluation	43
3.4	Les systèmes d'échanges et de partages de textes	44
3.4.1	Principes généraux	44
3.4.2	Activités associées	44
3.4.3	Interactions et feedbacks	46
3.4.4	Engagement des étudiants	47
3.4.5	Évaluation	48
3.5	Synthèse de l'étude	48
3.6	Les verrous identifiés	50

3.6.1	Le support de la saisie de texte libre	50
3.6.2	La fourniture de feedbacks textuels	51
3.6.3	Susciter la confrontation d'idées entre pairs	51
3.6.4	La mise en œuvre d'auto-évaluations et d'évaluations par les pairs	52
3.6.5	L'utilisation des tests sommatifs à visée formative	52
3.7	Conclusion	53

3.1 Introduction

L'étude proposée dans ce chapitre détermine dans quelle mesure les dispositifs existants participent efficacement à la mise en œuvre de l'évaluation formative telle qu'elle est décrite dans le modèle théorique du chapitre précédent. Cette étude vise l'identification des verrous qui subsistent à la bonne mise en œuvre d'évaluations formatives à l'aide de dispositifs technologiques dans les amphithéâtres d'universités.

Dans la suite du document, le modèle théorique d'évaluation formative décrit dans la section 2.6 sera désigné par l'acronyme MEF (Modèle d'Évaluation Formative).

3.2 La méthode d'analyse

3.2.1 Les dispositifs retenus dans l'étude

Nous avons retenu, dans notre étude, les dispositifs répondant à deux critères :

1. l'utilisabilité dans un amphithéâtre d'université dans le contexte de cours traditionnels ;
2. l'utilisabilité quel que soit le champ disciplinaire et la matière enseignée.

Les dispositifs inclus dans l'étude s'inscrivent donc dans les contraintes socio-économiques décrites en introduction : ils s'intègrent dans un cours en face à face traditionnel (format cours magistral ponctué par l'utilisation des dits dispositifs) pouvant s'adresser à de larges effectifs (au-delà de 30 étudiants). Typiquement les dispositifs peuvent être utilisés dans un amphithéâtre d'université en étant compatibles avec les modes d'enseignement pratiqués par un grand nombre d'enseignants. Ceux-ci disposent de beaucoup de matériaux réutilisables (supports

de cours, diaporamas, etc.) que les dispositifs étudiés ne doivent pas remettre en cause. Nos travaux ne se situent donc pas dans la mouvance de nouvelles modalités d'enseignement de type "classe inversée" ou "apprentissage par projet".

Les SVI sont déjà identifiés dans la littérature comme des dispositifs permettant la mise en œuvre d'évaluations formatives dans les amphithéâtres, ils rentrent donc directement dans le champ de notre étude.

D'autre part, l'arrivée dans les amphithéâtres de dispositifs connectés à Internet a rendu de nouvelles interactions possibles; les types d'interactions ne sont plus cantonnés à la fourniture d'un numéro identifiant un choix dans une question à choix multiples; les étudiants peuvent utiliser leur smartphone, leur tablette ou ordinateur portable pour échanger ou partager des documents ou messages textuels. Différentes situations pédagogiques ont émergé de l'utilisation de ces nouvelles possibilités, dont certaines telles que la prise de notes collaborative ou le micro-blogging respectent nos contraintes socio-économiques et s'inscrivent dans une ou plusieurs des stratégies du MEF.

Les dispositifs que nous avons retenus dans notre revue sont donc de deux natures :

- les systèmes de votes interactifs (SVI);
- les systèmes d'échanges et de partages de textes (SEPT).

3.2.2 La grille d'évaluation

La grille d'évaluation utilisée dans notre étude est synthétisée dans le Tableau 3. Cette grille présente les critères retenus pour qualifier les systèmes étudiés ainsi que les domaines de valeurs associés; ces critères et domaines de valeurs sont détaillés dans les paragraphes suivants.

Table 3. 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		ne participe pas
A-2	Le dispositif participe à l'activité de questionnement de la classe		participe
			réalise

Table 3. Grille de critères d'évaluation des dispositifs étudiés

Code	Description	Domaine de valeurs						
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							

Table 3. Grille de critères d'évaluation des dispositifs étudiés

Code	Description	Domaine de valeurs	
I-TF	Types de feedback fournis par le dispositif en sortie		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		engage moins d'un tiers des étudiants
			engage entre un tiers et deux tiers des étudiants
			engage plus de deux tiers des étudiants

3.2.2.1 Les critères

Les critères sont répartis en quatre catégories : la catégorie "Activités de référence" (code préfixé par "A-"), la catégorie "Stratégies" (code préfixé par S-), la catégorie "Interactions" (code préfixé par "I-") et la catégorie "Engagement" (code préfixé par E-).

Les critères de la catégorie "Activités de référence" sont les critères permettant de qualifier les dispositifs étudiés vis-à-vis des cinq activités de référence décrites dans le MEF. Il est indispensable de pouvoir qualifier les systèmes étudiés vis-à-vis de ces activités, cependant cette qualification n'est pas suffisante. En effet, certaines activités supportées par les systèmes étudiés sont différentes des activités de référence, mais elles s'insèrent clairement dans une des stratégies caractérisant l'évaluation formative. Il convient donc de qualifier également les dispositifs au

regard des cinq stratégies du MEF. Les critères nécessaires à cette qualification sont contenus dans la catégorie "Stratégies". La stratégie S2 est découpée en deux critères distincts (S-2.1 et S-2.2) afin de qualifier plus finement les dispositifs. En effet, nous avons constaté que des dispositifs peuvent participer effectivement à la conception de situations efficaces de discussions en classe sans pour autant éliciter des preuves de performances des étudiants sur leur travail.

La catégorie "Interactions" s'intéresse aux critères permettant de qualifier comment un dispositif interagit avec les autres agents dans le contexte des différentes activités supportées. Le critère I-ST permet de qualifier les dispositifs étudiés en fonction des types d'interactions proposés en entrée : par exemple les systèmes de votes interactifs permettent en général des interactions très limitées comme le choix d'une réponse parmi N, alors que les systèmes de prises de notes collaboratifs permettent la saisie de texte libre. Le type d'interaction impacte la nature des actions réalisées par les étudiants (rédiger un texte est plus complexe que de cliquer sur un bouton désignant une réponse), mais aussi la nature du feedback quand le résultat des interactions est partagé (un texte partagé s'apparente à un feedback sous forme de commentaire textuel, alors que le choix d'une réponse par les étudiants ne sera retranscrit que sous la forme d'un nombre représentant le pourcentage de réponses correspondant à ce choix). Le critère I-TF permet de qualifier les dispositifs étudiés en fonction des types de feedback restitués : textuels ou uniquement sous forme d'indicateurs numériques (scores, pourcentages, échelles de Lickert). En effet, comme indiqué dans le paragraphe 2.3.2, les feedbacks prenant la forme de commentaires textuels s'avèrent plus efficaces que ceux prenant la forme d'indicateurs numériques.

Enfin la catégorie "Engagement" (code préfixé par E-) s'intéresse à l'impact du dispositif sur les apprenants en terme d'engagement. Le critère E-TE permet de qualifier les dispositifs étudiés sous l'angle de leur capacité à engager un grand nombre d'étudiants. En effet, un dispositif peut participer à différentes stratégies de mise en œuvre d'évaluation formative tout en ne concernant qu'un nombre restreint d'étudiants. Notons que Fredricks *et al.* (2004) rappellent que trois types d'engagement ont été identifiés (comportemental, émotionnel, et cognitif), et qu'il est recommandé d'étudier ce concept sous ses différentes facettes. Dans le contexte de nos travaux, nous nous sommes intéressés uniquement à l'engagement comportemental, plus précisément à l'engagement en terme de participation à des activités proposées durant les cours en face à face où les étudiants n'ont pas obligation de participer. Nous émettons l'hypothèse que l'engagement

comportemental sur des activités complexes et optionnelles dépasse le cadre comportemental, cependant, durant nos expérimentations, nous n'avons pu mesurer de manière objective que les taux de participation des étudiants aux activités proposées.

3.2.2.2 Les domaines de valeurs

Le domaine de valeurs associé aux critères des différentes activités de référence est une échelle de Lickert à trois niveaux :

- "ne participe pas" ;
- "participe" ;
- "réalise" ;

Le premier niveau "ne participe pas" signifie que le dispositif ne s'inscrit en rien dans l'activité concernée. Il n'assiste pas la situation d'apprentissage pour la mise en œuvre de l'activité. Le deuxième niveau "participe" indique que le dispositif contribue à la mise en place de l'activité. Par exemple, les SVI contribuent à mettre en place une situation d'évaluation sous forme de commentaires quand l'enseignant commente les résultats affichés par le SVI. Pour autant, les SVI ne sont pas utilisés pour réaliser ces commentaires. Enfin le troisième niveau "réalise" indique que le dispositif est l'élément par lequel l'activité est réalisée ; par exemple un SVI est un outil permettant de questionner une classe.

Le domaine de valeurs associé aux critères des différentes stratégies du MEF est la même échelle de Lickert à trois niveaux que celle utilisée pour les critères de la catégorie "Activités de référence". Le premier niveau "ne participe pas" signifie que le dispositif ne s'inscrit en rien dans la stratégie concernée. Il n'assiste pas la situation d'apprentissage dans le sens de la stratégie. Le deuxième niveau "participe" indique que le dispositif contribue à la mise en place de la stratégie. Par exemple, les SVI contribuent à mettre en place une situation de discussions entre étudiants s'ils sont utilisés en support d'une approche "peer instruction" (Crouch et Mazur, 2001). Pour autant, ces systèmes ne sont pas utilisés pour réaliser cette discussion. Enfin le troisième niveau "réalise" indique que le dispositif est un élément qui permet de réaliser la stratégie ; par exemple un système de votes interactifs est bien le fournisseur de feedback sous forme de pourcentages. Il n'est pas un outil qui facilite la mise à disposition d'un tel feedback, il est bien l'agent qui réalise le feedback.

Cette approche distinguant la notion de participation de la notion de réalisation permet de manière assez simple de représenter le degré de responsabilité de l'agent technologique dans chacune des activités ou stratégies. L'approche a été directement inspirée des matrices de responsabilités de type RACI¹ utilisées pour clarifier les rôles et responsabilités des acteurs au sein d'un projet.

Le critère I-ST quand à lui s'évalue de trois manières : le dispositif ne propose pas la saisie de texte, il propose la saisie de texte court (un mot ou une expression) ou il propose la saisie de texte libre.

Le critère I-TF s'évalue aussi suivant trois valeurs : le dispositif ne fournit pas de feedback, il fournit un feedback sous forme d'indicateurs numériques, ou il fournit un feedback sous forme textuelle.

Enfin le domaine de valeur associé au critère E-TE est une échelle de Lickert à trois niveaux inspirée de l'échelle utilisée par Mazur (1997) dans son usage des systèmes de votes interactifs ; l'échelle proposée par Mazur a été adaptée pour évaluer le niveau d'engagement que suscite le dispositif auprès des étudiants :

- "engage moins d'un tiers des étudiants" ;
- "engage entre un et deux tiers des étudiants" ;
- "engage plus de deux tiers des étudiants" ;

Comme indiqué dans le Tableau 3, chaque valeur de chaque domaine est représentée par une couleur (blanc, gris ou noir) ; cette approche vise l'obtention d'une synthèse graphique du degré d'implication et d'efficacité des dispositifs pour la mise en œuvre d'évaluation formative.

3.2.3 Le processus d'analyse

Pour chaque dispositif, l'étude a consisté à rassembler les éléments permettant de qualifier le dispositif à l'aide des critères des différentes catégories. Nous nous sommes donc intéressés aux activités pour lesquelles chaque dispositif était utilisé dans le contexte de cours en face-à-face. Cet axe de lecture nous a permis de renseigner les critères des catégories "Activités de référence" et "Stratégies". L'étude des principales fonctionnalités de chaque dispositif sous l'angle des interactions et feedbacks proposés nous a permis de renseigner les critères de la catégorie "Interactions". Enfin, la recherche des données chiffrées sur le taux d'engagement des étudiants dans les expérimentations étudiées nous a permis de renseigner le

1. https://en.wikipedia.org/wiki/Responsibility_assignment_matrix

critère de la catégorie "Engagement".

Les résultats sont donc présentés pour chaque type de dispositifs selon la même disposition : (1) Principes généraux, (2) Activités associées, (3) Interactions et feedbacks, (4) Engagement des étudiants, et (5) Évaluation.

La prochaine section présente les résultats associés aux systèmes de votes interactifs.

3.3 Les Systèmes de votes interactifs

3.3.1 Principes généraux

Les Systèmes de Votes Interactifs (SVI) permettent à une classe entière ou un amphithéâtre de répondre à une question à choix multiples affichée sur un écran (Kay et 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 (Uhari *et al.*, 2003; Gauci *et al.*, 2009) ainsi que pour améliorer les résultats des étudiants participant aux votes interactifs (Shaffer et Collura, 2009). L'étude de Gauci *et al.* (2009) 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 socio-cognitifs.

Les SVI se sont imposés comme outils de référence pour mettre en œuvre l'évaluation formative dans les classes ou amphithéâtres. L'étude détaillée des activités pédagogiques associées aux SVI présentée dans le paragraphe suivant confirme cet état de fait en mettant en lumière les différentes activités de références et stratégies du MEF supportées par les SVI.

3.3.2 Activités associées

Les SVI s'utilisent dans la classe ou dans l'amphithéâtre. Ils permettent de ponctuer le déroulement du cours par une liste de questions posées à l'assemblée dont les résultats récoltés sont présentés à l'ensemble des étudiants. La littérature scientifique autour des SVI fait apparaître différentes modalités de poses de questions et des articulations des questions avec le cours qui varient d'une mise en œuvre à l'autre. En synthèse, quatre situations pédagogiques ont été clairement identifiées.

3.3.2.1 La situation nominale

Ce que nous qualifions de situation nominale est la situation correspondant au schéma d'utilisation le plus simple :

1. l'enseignant présente une question à l'assemblée à l'aide d'un vidéo-projecteur ;
2. les étudiants répondent à la question à l'aide de leur dispositif (boîtier, smartphone, tablette, etc.) ;
3. les résultats sont présentés sous forme graphique à l'assemblée (histogramme, diagramme circulaire, etc.) ;
4. une discussion entre pairs, entre enseignant et étudiants suit la présentation des résultats.

Cette approche est décrite dans diverses publications pointant les bénéfices des SVI (Burnstein et Lederman, 2001; Uhari *et al.*, 2003; Elliott, 2003; Draper et Brown, 2004; Cline, 2006; Cutts, 2006; Trees et Jackson, 2007). Les questions sont des questions à choix (multiple, exclusif ou vrai/faux). Les résultats présentés sous forme de diagrammes indiquent la répartition des choix fournis par les étudiants. Les résultats sont présentés de manière anonyme.

Le questionnement de l'assemblée et le recueil des réponses sont effectués à l'aide du SVI (critère A-2). Les résultats affichés à l'assemblée correspondent à un feedback numérique adressé à tous les étudiants et fourni par le dispositif (critère S-3). Les informations fournies sont des preuves du niveau de connaissance des étudiants (critère S-2.2) permettant à chacun de positionner son niveau de connaissance par rapport au sujet abordé de manière absolue (l'étudiant sait s'il a répondu de manière correcte ou incorrecte à la question) mais aussi de manière relative par rapport à l'ensemble des étudiants ayant répondu (critère S-5).

La discussion déclenchée après la présentation des résultats permet d'apporter des explications complémentaires ou de résoudre des conflits socio-cognitifs (critère S-2.1). Ces explications peuvent être apportées par l'enseignant ou par les pairs (critère S-4).

Dans leur utilisation la plus basique, les SVI réalisent une des cinq activités de références et s'inscrivent dans quatre des cinq stratégies du MEF. Très rapidement, des approches d'utilisations différentes ont été proposées et ont renforcé les atouts de l'approche nominale. Parmi ces approches alternatives, celle désignée par l'expression "Instruction par les pairs" remporte encore aujourd'hui un franc succès.

3.3.2.2 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, à sonder les étudiants 2 fois : une première fois après une phase de réflexion individuelle, puis une deuxième fois après que les étudiants ont été invités à convaincre leurs voisins du bien fondé de la réponse qu'ils ont fournie à l'issue du premier sondage. Les résultats du deuxième sondage révèlent d'une part un plus grand nombre de bonnes réponses mais surtout les bénéfices en terme d'apprentissage se trouvent renforcés par rapport à une approche classique (Mazur, 1997). Mazur introduit par la même occasion la notion de *ConceptTests* correspondant à un lot de questions permettant de tester la compréhension d'un sujet plus que la connaissance brute.

Devant les résultats probants apportés par l'approche IP (Crouch et Mazur, 2001) beaucoup d'universités se sont lancées dans l'utilisation des SVI en appliquant la méthode de Mazur (Burnstein et Lederman, 2001; Cline, 2006; Draper et Brown, 2004; Trees et Jackson, 2007).

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 (critère S-2.1), et d'autre part le fait que les étudiants deviennent des ressources d'instructions pour les autres (critère S-4).

3.3.2.3 L'approche TEFA

Beatty et Gerace (2009) décrivent l'approche pédagogique TEFA (Technology Enhanced Formative Assessment) comme une approche pour enseigner les sciences

en s'appuyant sur des SVI. La méthode TEFA s'inscrit dans une longue tradition de recherches et d'expérimentations menées à l'Université du Massachusetts. Dufresne *et al.* (1996) présentent l'utilisation d'un des premiers SVI commercialisés, *Classtalk*, en proposant déjà une approche pédagogique dédiée. Cette approche fera l'objet d'une première formalisation, la méthode "assessing to learn" (Dufresne et Gerace, 2004), qui sera reprise et affinée en 2009 pour devenir la méthode TEFA.

TEFA repose sur quatre principes :

1. motiver et concentrer l'apprentissage des étudiants à l'aide d'une *instruction guidée par les questions* ;
2. développer la compréhension et la maîtrise scientifique des étudiants à l'aide de *discours dialogiques* ;
3. éclairer et ajuster l'enseignement et l'apprentissage à l'aide d'évaluations formatives ;
4. aider les étudiants à développer des compétences méta-cognitives et à coopérer au processus d'apprentissage.

Centrée sur les questions, TEFA se déroule au rythme du "cycle de question" répété autant de fois que possible durant une session d'apprentissage :

1. L'enseignant pose une question ou un problème aux étudiants.
2. Les étudiants réfléchissent à la question, seuls et/ou en petits groupes et décident d'une réponse.
3. Un SVI est utilisé pour collecter les réponses et celles-ci sont restituées à l'aide d'un diagramme.
4. Les étudiants sont invités à justifier le choix de leur réponse sans révéler quelle réponse est la bonne ; dans la pratique, cette étape et la précédente s'entremêlent.
5. L'enseignant anime la discussion entre pairs avec pour objectif l'accroissement de la compréhension de l'ensemble.
6. L'enseignant propose une synthèse clôturant le cycle.

L'approche TEFA se distingue de l'approche IP principalement en raison de son premier principe. En effet, dans l'approche TEFA, la totalité du cours se construit sur les questions : les questions sont au cœur du dispositif de construction des connaissances, là où l'approche IP utilise les *ConceptTests* pour contrôler le niveau de

compréhension d'une notion abordée à travers des dispositifs traditionnels (lecture, présentation magistrale) (Beatty et Gerace, 2009). Beatty et Gerace revendiquent l'appartenance de TEFA à la mouvance des approches d'apprentissage par projet ou d'apprentissage à base de problèmes.

Le cycle de question restant cependant très proche du processus décrit par l'IP, les mêmes stratégies d'évaluation formative se retrouvent renforcées (critères S-2.1 et S-4).

3.3.2.4 L'approche par "Séquence de questions"

Reay *et al.* (2008) décrivent une approche consistant à proposer une série de questions sur un même sujet. Les auteurs font l'hypothèse que le traitement d'un même concept à l'aide d'une séquence de questions soigneusement élaborée peut accroître le niveau de compréhension. Deux types de séquences ont été modélisées : les séquences de type "facile-difficile-difficile" et les séquences de type "tir rapide".

Dans le cas d'une séquence de type "tir rapide", les différentes questions sont plutôt d'un faible niveau de difficulté. Elles permettent aux étudiants d'exercer leurs compétences sur des problèmes simples mais dans différents contextes.

Dans le cas d'une séquence de type "facile-difficile-difficile", l'objectif est d'amener les étudiants à exercer leur compétence sur des questions de difficulté croissante en commençant par une question facile afin de les mettre en confiance et donc de susciter un fort engagement sur les questions les plus difficiles. L'approche par "Séquence de questions" peut être utilisée conjointement avec l'approche IP.

L'étude de Reay *et al.* (2008) n'effectue pas de comparatif entre un cours faisant usage d'une séquence de questions et un cours n'utilisant qu'une question par concept. Néanmoins, l'étude confirme l'efficacité du dispositif comparé à la non utilisation de SVI.

Cette approche, en intensifiant le nombre de questions posées aux étudiants (critère A-2) sur un même concept, augmente l'opportunité de feedbacks adressés aux étudiants sur un sujet donné (critères S-2.2 et S-3).

3.3.3 Interactions et feedbacks

Toutes les revues scientifiques (Abrahamson, 1999; Beatty, 2005; Fies et Marshall, 2006; Caldwell, 2007; Kay et LeSage, 2009) présentent les SVI comme des dispositifs permettant à une assemblée d'étudiants de répondre à des questions

à choix multiples. Les questions à choix offrent un avantage évident : il est facile de restituer une synthèse des résultats obtenus compréhensible par tous en un temps minimum.

Comme nous l'avons déjà mentionné dans le paragraphe 3.2.1, l'arrivée dans les amphithéâtres de dispositifs connectés à Internet a rendu possible de nouvelles interactions. Nous n'avons recensé qu'une seule étude présentant des travaux visant à profiter de la nature des nouveaux dispositifs utilisés par les étudiants pour améliorer les types d'interactions au sein d'un SVI. McLoone *et al.* (2015) proposent un SVI entièrement logiciel permettant aux étudiants de saisir une réponse sous la forme d'une entrée complètement libre. Les étudiants peuvent donc saisir un schéma, une formule complexe ou tout autre texte libre en réponse à une question posée par l'enseignant. Dans ce système baptisé SDSRS (Smart Device Student Response System), la présentation des résultats à l'ensemble des apprenants consiste soit (1) à présenter l'ensemble des réponses à toute l'assemblée soit (2) à présenter un sous-ensemble de réponses sélectionnées par l'enseignant en temps réel. Un tel dispositif ne peut pas s'appliquer à une large audience : l'affichage de 50 schémas dans un amphithéâtre ne correspond pas une restitution recevable de manière efficace ; l'évaluation d'une cinquantaine de schémas par un enseignant en temps réel ne semble pas non plus être une solution soutenable.

Le SVI Clica (YAMAMOTO *et al.*, 2014) tire parti lui aussi d'interactions plus riches comme la saisie de texte libre, mais ce type d'interaction n'est pas proposé pour répondre au sondage ; en effet, Clica propose une fonctionnalité d'échanges de message textuels afin de fournir un moyen alternatif d'interaction entre les membres d'un large groupe d'étudiants.

Plusieurs SVI commerciaux² permettent la saisie de texte pour la fourniture de réponses courtes. Les résultats sont alors affichés sous forme de diagramme ou de nuage de mots dans lequel les réponses les plus fréquemment données apparaissent dans une taille de police supérieure. Nous n'avons cependant trouvé aucune étude scientifique sur les SVI disposant de ces interactions.

En synthèse, les SVI destinés aux larges audiences proposent des types d'interactions restreints permettant de répondre à des questions à choix ou à des questions à réponse courte. Les études sont focalisées sur l'utilisation de questions à choix. Les feedbacks restitués sont des feedbacks correspondant à des

2. <https://www.turningtechnologies.com>, <https://www.polleverywhere.com>, <http://qwizdom.com/>

indicateurs numériques (pourcentage obtenu pour chaque réponse) restitués sous forme graphique (diagrammes ou nuages de mots).

3.3.4 Engagement des étudiants

La totalité des articles déjà mentionnés dans notre revue indique les bénéfices apportés par les SVI en terme de capacité à engager les étudiants. Nous nous sommes intéressés plus particulièrement aux études ayant mesuré de manière objective le taux de participation des étudiants à l'activité de réponse aux questions posées via un SVI. Les résultats obtenus présentent des taux de participation supérieurs à 80% (Greer et Heaney, 2004) et majoritairement proche des 90% (Bullock *et al.*, 2002; Caldwell, 2007; Uhari *et al.*, 2003).

Dans la plupart des études, le caractère anonyme des modalités de participation, d'évaluation et de présentation des résultats est présenté comme jouant un rôle décisif dans ces taux de participation élevés (Kay et LeSage, 2009).

3.3.5 Évaluation

Le Tableau 4 propose une synthèse de l'évaluation des SVI suivant les critères de la grille présentée dans le Tableau 3.

Les SVI permettent de réaliser une des cinq activités de référence : l'activité de questionnement de la classe (critère A-2). Les SVI participent à toutes les stratégies concernées par la mise en œuvre d'évaluations formatives. Les SVI réalisent directement certaines de ces stratégies : l'apport de preuve du niveau d'apprentissage des étudiants sur un thème donné (critère S-2.2) et la fourniture d'un feedback permettant aux étudiants d'orienter leurs efforts (critère S-3). Bien qu'aucune étude ne mentionne explicitement le fait que les SVI participent à la clarification et au partage des objectifs d'apprentissage et des critères de succès (critères A-1 et S-1), beaucoup d'entre elles indiquent que l'utilisation de tels systèmes améliorent sensiblement les résultats des étudiants aux examens terminaux (Brewer, 2004; Bullock *et al.*, 2002; Caldwell, 2007; Crouch et Mazur, 2001; Kennedy et Cutts, 2005). Nous avons considéré que ces résultats inscrivaient de facto les SVI comme systèmes participant à la clarification et au partage des objectifs d'apprentissage et des critères de succès.

La nature des interactions et des feedbacks proposés est relativement basique : les interactions proposées sont essentiellement destinées à permettre de répondre

Table 4. Synthèse de l'évaluation des systèmes de votes interactifs

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

à des questions à choix (critère I-ST); les feedbacks (critère I-TF) se présentent en conséquence sous forme d'indicateurs numériques (pourcentage de réponses correspondant à chacun des choix possibles).

En terme d'engagement, les SVI sont très mobilisateurs (critère E-TE) : plus de 80% des étudiants participent aux activités dispensées via les SVI.

3.4 Les systèmes d'échanges et de partages de textes

3.4.1 Principes généraux

Les systèmes d'échanges et de partages de textes (SEPT) permettent aux étudiants participant à un même cours en face à face d'échanger ou de partager des messages textuels directement en relation avec le cours. Ces messages peuvent être des notes ou des commentaires relatifs au cours ou des questions que les étudiants se posent les uns aux autres. La littérature scientifique fait émerger fortement deux situations mobilisant les SEPT en amphithéâtre : (1) la prise de notes collaborative pratiquée sur un document électronique partagé, et (2) la prise de notes collaborative pratiquée via une approche de micro-blogging.

3.4.2 Activités associées

3.4.2.1 La prise de notes dans un document unique

La prise de notes est une activité importante des étudiants participant à un cours. Dans leur étude (Steimle *et al.*, 2007), les auteurs montrent que la proportion d'étudiants prenant des notes pendant le cours varie entre 60% et 95%. Toutes les disciplines sont concernées. En conséquence, la prise de notes est une activité durant laquelle les étudiants produisent du contenu en réaction au cours auquel ils participent et ceci quel que soit la discipline ou la matière enseignée.

L'étude de Kam *et al.* (2005) présente les résultats et bénéfices de l'utilisation de l'outil de prise de notes collaborative Livenotes. Le partage, la confrontation des différentes contributions et l'exploitation des contributions pendant le cours par l'enseignant sont autant de facteurs positifs pour l'apprentissage déclenchés par la prise de notes collaborative.

Dans leur étude sur les différentes modalités de prise de notes, Kiewra *et al.* (1988) ont montré que les étudiants atteignent un niveau de rétention maximum des connaissances diffusées pendant le cours quand ils peuvent annoter les supports. Les systèmes décrits dans Kam *et al.* (2005), Simon *et al.* (2008) ou Singh *et al.* (2004) offrent la possibilité d'importer la diapositive présentée pendant le cours dans l'interface donnée aux étudiants pour prendre des notes.

L'essor d'Internet a fait émerger différents systèmes de prises de notes collaborative utilisables avec n'importe quel dispositif connecté à Internet. L'utilisation d'un wiki pour la prise de notes collaborative pendant les cours présentée dans O'Neill (2005) ou l'utilisation d'un éditeur de texte collaboratif décrit dans Habert (2012) confirment les différents bénéfices de la prise de notes collaborative. Sur le volet non-académique, la société Unishared³ fournit une plate-forme collaborative de prise de notes reposant sur Google Drive.

En synthèse, les systèmes de prises de notes collaboratives permettent aux étudiants de mutualiser leurs efforts de prise de notes qui deviennent ainsi des ressources d'instructions pour les autres (critère S-4). Quand le document produit par les étudiants est exploité par l'enseignant pendant le cours, il offre l'opportunité à l'enseignant de fournir du feedback aux étudiants (Kam *et al.*, 2005) (critère S-3).

3.4.2.2 Le partage de N micro-documents individuels

Le Web 2.0 et les réseaux sociaux remettent la production individuelle au cœur des modalités de communication et de partage. Les individus ne contribuent pas à une œuvre commune, ils partagent leurs contributions individuelles. Cette approche est déjà décrite dans Singh *et al.* (2004) où les étudiants prennent leurs notes à l'aide d'un assistant numérique personnel (PDA). Les notes prises par chaque étudiant sont alors envoyées aux autres étudiants qui peuvent les réutiliser pour constituer leurs propres notes. Cette approche présage déjà de l'utilisation du micro-blogging dans les amphithéâtres.

3. <http://www.unishared.com>

Spence (2011) définit un micro-blog comme un weblog dans lequel chaque article est limité à 140 caractères et auquel s'ajoutent des fonctionnalités de réseau social. Bien que plusieurs applications web supportent le micro-blogging telles que Indenti.ca⁴ ou Plurk⁵, c'est avec la plateforme Twitter que le micro-blogging s'est rapidement imposé comme un outil efficace de partage et de construction collaborative de connaissances pouvant être utilisé à tout moment, n'importe où, avec tous les périphériques connectés à Internet (Cheng *et al.*, 2009).

Différentes utilisations du micro-blogging ont été expérimentées dans l'éducation. Plusieurs études (Borau *et al.*, 2009; Ebner *et al.*, 2010; Grosseck et Holotescu, 2008; Reinhardt *et al.*, 2009) montrent les bénéfices que le micro-blogging peut offrir en tant qu'activité d'apprentissage. Reinhardt *et al.* (2009) indiquent notamment que le micro-blogging est un moyen de fournir du feedback durant un cours en face à face : un feedback adressé à l'enseignant mais aussi du feedback entre pairs que se procurent les contributeurs (critères S-3 et S-4).

Les messages postés à l'aide d'un dispositif de micro-blogging peuvent être des réponses à des messages postés précédemment. Ces fonctionnalités permettent de déclencher de véritables discussions pendant le cours ou en dehors du cours (critère S-2.1).

Enfin, l'approche micro-blogging offre des fonctionnalités de personnalisation avancées permettant à chaque étudiant de personnaliser des vues sur un grand nombre de contributions. Ainsi, chaque étudiant est capable de retenir ou de rejeter les contributions en utilisant les fonctionnalités de recherche à facette s'appuyant sur des *hashtags*, mentions et favoris. Ces fonctionnalités de personnalisation participent au fait que les apprenants s'approprient leur apprentissage (critère S5).

3.4.3 Interactions et feedbacks

Par définition, les systèmes d'échange et de partage de textes (SEPT) permettent la saisie de texte libre ; en conséquence, ces textes partagés deviennent des feedbacks fournis sous forme textuelle.

4. <https://identi.ca/>

5. <http://www.plurk.com/>

3.4.4 Engagement des étudiants

Concernant l'approche de partage d'un document unique, chaque étudiant doit écrire sur le même document pour collaborer. Même si cette approche est techniquement viable aujourd'hui, elle ne facilite pas la participation d'un grand nombre d'étudiants. Les systèmes présentés dans Kam *et al.* (2005) ou Singh *et al.* (2004) sont conçus pour travailler avec de petits groupes de cinq étudiants au plus. Dans Delangue (2012), Clément Delangue, un des trois étudiants cofondateurs d'Unishared, explique que, dans la pratique, afin d'éviter le désordre dans le document partagé, seuls deux à trois étudiants contribuent à la rédaction du document pendant le cours. En conséquence, pour les cours s'adressant à un large auditoire, cette approche ne favorise pas la participation du plus grand nombre. De fait, certains des avantages de la prise de notes collaborative sont réduits : les retours obtenus par les enseignants ne concernent qu'une très petite partie de l'auditoire et il n'y a aucun moyen de savoir quels sont les réels bénéfices pour les étudiants qui ne prennent pas part à la prise de notes collaborative. Ces éléments nous incitent à classer les systèmes de partage d'un document unique dans la catégorie des systèmes à faible capacité d'engagement des étudiants.

L'approche de partage de N micro-documents est incarnée principalement par le micro-blogging. Les plateformes comme Twitter ont été conçues sans limites théoriques sur le nombre de contributeurs; en d'autres termes, tous les étudiants d'un amphithéâtre aussi grand et rempli qu'il soit peuvent participer aux échanges diffusés via la plateforme de micro-blogging. Quand l'utilisation du micro-blogging fait partie d'une activité préparée par l'enseignant, les taux de participation affichés sont proches des 90% (Borau *et al.*, 2009; Ebner *et al.*, 2010). Nous n'avons malheureusement trouvé aucune étude mentionnant les taux de participation quand les étudiants sont invités à utiliser le micro-blogging sur la base du volontariat. Il nous semble donc que l'évaluation la plus juste du taux participation dans un tel contexte (contexte de volontariat correspondant au contexte dans lequel nous avons pu évaluer les taux de participation des systèmes de votes interactifs) consiste à s'appuyer sur les taux de contribution constatés sur Twitter. En avril 2014, le Wall Street Journal publie les résultats d'une étude indiquant que 44% des titulaires d'un compte Twitter n'ont jamais envoyé un message (Koh, 2014). Cet indicateur nous incite à classer les systèmes de partage de N micro-documents dans la catégorie des systèmes à capacité d'engagement moyenne.

Table 5. Synthèse de l'évaluation des systèmes d'échange et de partage de textes

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

3.4.5 Évaluation

Le Tableau 5 propose une synthèse de l'évaluation des systèmes d'échange et de partage de textes (SEPT) suivant les critères de la grille présentée dans le Tableau 3.

Les SEPT ne participent à aucune des activités de référence, ce qui explique sans doute le fait qu'ils ne soient pas référencés explicitement dans les systèmes permettant la mise en œuvre d'évaluation formative. Cependant, notre étude met en lumière le fait qu'ils participent à quatre stratégies relevant de l'évaluation formative. En effet, ces systèmes permettent aux étudiants de discuter entre eux dans l'amphithéâtre (critère S-2.1) comme à l'extérieur de l'amphithéâtre en s'échangeant des messages textuels (I-ST). Ils sont un moyen de fournir du feedback par les pairs (critère S-3 et S4) restitué sous forme textuelle (critère I-TF). Les fonctionnalités de personnalisation des SEPT de type micro-blogging favorisent l'appropriation par les étudiants de leur propre apprentissage (critère S-5).

Néanmoins, les SEPT n'apportent aucune réponse en terme de clarification des objectifs d'apprentissage et des clés de succès (critère S-1); d'autre part, ils n'exhibent aucune preuve objective de l'avancement des étudiants dans leur apprentissage pouvant être traitée en temps réel. Leur capacité à engager un grand nombre d'étudiants est, dans le meilleur des cas, moyenne (critère E-TE).

3.5 Synthèse de l'étude

Le Tableau 6 présente une synthèse des évaluations des systèmes étudiés dans cette revue de la littérature. L'évaluation des systèmes de vote interactifs montre que les SVI sont résolument des systèmes permettant la mise en œuvre d'évaluation formative dans les amphithéâtres d'université en engageant un grand nombre d'étudiants. Cependant, certaines activités de référence ne sont pas supportées par les SVI : l'auto-évaluation et l'évaluation par les pairs, pas plus que l'utilisation de tests sommatifs à visée formative. D'autre part, certaines des stratégies

Table 6. Synthèse de l'évaluation des systèmes étudiés

	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	E-TE	
SVI																
SEPT																
Systeme cible																

dans lesquelles s'inscrivent les SVI (clarification des objectifs, déclenchement de discussions, utilisation des pairs comme ressources d'apprentissage) ne sont pas directement réalisées par ces systèmes : ils agissent uniquement comme supports aux activités que pourrait organiser l'enseignant pour mettre en œuvre ces stratégies. Enfin, les types d'interactions restreints mènent à des feedback restitués sous forme d'indicateurs numériques uniquement.

L'évaluation des systèmes d'échange et de partage de textes comblent des lacunes des SVI : les types d'interactions et d'échanges de textes libres conduisent à des feedbacks fournis sous forme textuelle. Ces outils permettent directement aux étudiants de discuter et réalisent de fait l'utilisation des pairs comme ressource d'instruction. Ces outils n'offrent cependant pas l'impact des SVI en terme de participation, ils ne supportent aucune des activités de référence et n'investissent pas certaines stratégies décisives pour relever entièrement de l'évaluation formative (pas de preuve d'avancement des étudiants dans leur apprentissage, pas de participation à la clarification des objectifs).

La dernière ligne du Tableau 6 synthétise les caractéristiques de notre système cible : un système intégrant l'ensemble des activités de référence et combinant les qualités des SVI avec celles des SEPT pour aboutir à un système d'évaluation formative utilisable dans les amphithéâtres d'universités surpassant les SVI. Le paragraphe suivant présente les verrous à lever pour concevoir et réaliser un tel système.

3.6 Les verrous identifiés

3.6.1 Le support de la saisie de texte libre

Nous avons identifié deux verrous principaux dans l'intégration de la saisie de texte libre dans le système cible.

3.6.1.1 L'évaluation en temps réel de réponses textuelles

Les SVI sont capables d'évaluer de manière formelle les réponses des étudiants en temps réel et de restituer immédiatement des résultats de manière compréhensible par tous. Ces traitements en temps réel sont possibles en raison de la simplicité des algorithmes mobilisés (décompte des réponses associées à chaque choix et affichage de diagrammes). L'évaluation de réponses fournies sous la forme de textes est plus

complexe et nécessite donc une révision du processus d'évaluation et des algorithmes utilisés pour l'évaluation et la restitution des résultats. L'intégration de saisie de texte libre soulève donc un premier lot de questions :

Comment, dans le système cible, évaluer de manière formelle en temps réel des réponses fournies sous forme de texte libre? Comment restituer les résultats de l'évaluation dans un amphithéâtre de manière compréhensible par tous?

3.6.1.2 Le maintien du haut niveau d'engagement

Les résultats de l'étude placent les SEPT dans la catégorie des systèmes à taux de participation moyen alors que les SVI sont capables de mobiliser un très grand nombre de participants. Faire participer les étudiants à des tâches d'écriture semble donc plus complexe que de les faire participer à un sondage nécessitant uniquement la sélection de choix. L'intégration de la saisie de texte libre pose donc le problème suivant :

Comment maintenir le taux de participation des étudiants au niveau constaté dans l'utilisation des SVI lorsqu'on exige des étudiants des réponses sous forme de texte libre?

3.6.2 La fourniture de feedbacks textuels

Les SVI ne fournissent pas de feedbacks sous forme textuelle. Les SEPT s'avèrent capables de déclencher la production de feedbacks provenant des pairs. Cependant ces feedbacks ne s'intègrent pas dans un dispositif d'évaluation formel. Le problème suivant se pose donc à nous :

Comment fournir des feedbacks textuels dans le contexte d'une évaluation formelle d'un groupe d'étudiants? Plus précisément, comment tirer parti, dans le système cible, des systèmes d'échange et de partage de textes pour la délivrance de feedbacks textuels dans le contexte d'une évaluation formelle d'un groupe d'étudiants?

3.6.3 Susciter la confrontation d'idées entre pairs

Les SEPT permettent la confrontation d'idées entre pairs en exposant à tous les contributions de chacun. Cette confrontation d'idées n'est cependant pas

mise en œuvre dans un processus d'évaluation. Les approches "instruction par les pairs" (Mazur, 1997) et TEFA (Beatty et Gerace, 2009) ont démontré les bénéfices de la phase de discussion, d'échanges entre pairs et de confrontation d'idées dans l'utilisation des SVI (Crouch et Mazur, 2001; Kay et LeSage, 2009). Cette confrontation de points de vue n'a lieu que si l'enseignant décide d'inscrire l'utilisation du SVI dans un processus qui prévoit ce temps de discussion. Il semble donc pertinent de prévoir un système capable d'instancier un processus prévoyant explicitement cette phase de confrontation d'idée. Ceci nous amène à une nouvelle problématique :

Quel processus intégrant explicitement la confrontation d'idées entre pairs implanter dans le système cible? Comment implanter ce nouveau processus?

3.6.4 La mise en œuvre d'auto-évaluations et d'évaluations par les pairs

Aucun des dispositifs étudiés (SVI ou SEPT) ne propose ou ne s'intègre dans un processus incluant les activités d'auto-évaluation ou d'évaluation par les pairs. Or ces activités font partie des activités recensées par Black et Wiliam (2009) dans la liste des activités constituant les processus d'évaluation formative. Ce double constat soulève deux questions :

Quel processus intégrant les activités d'auto-évaluation et d'évaluation par les pairs implanter dans le système cible? Comment implanter ce nouveau processus?

3.6.5 L'utilisation des tests sommatifs à visée formative

Les SEPT, par nature, ne s'intègrent pas dans un processus incluant l'utilisation de tests sommatifs à visée formative. Les SVI, quant à eux, sont agnostiques vis-à-vis de cette activité : en effet, il revient à l'enseignant de faire le choix de poser des questions pendant le cours qui seront ou non intégrées dans les tests sommatifs. Cependant, cette activité fait partie des activités recensées par Black et Wiliam (2009) dans la liste des activités constituant les processus d'évaluation formative. Dans un souci de conservation de toute liberté pédagogique de l'enseignant, la question soulevée par cet état de fait peut s'exprimer ainsi :

Comment *faciliter* dans le système cible l'utilisation de tests sommatifs à visée formative ?

3.7 Conclusion

L'étude des systèmes existants participant à la mise en œuvre de l'évaluation formative durant les cours en face-à-face a mis en lumière un ensemble de verrous à lever pour la conception de systèmes supportant davantage d'activités et de stratégies caractéristiques de l'évaluation formative. La suite du manuscrit détaille la manière dont nous avons levé ces verrous, ainsi que les résultats obtenus.

Deuxième partie

Systeme d'évaluation d'audience

Chapitre 4

Modélisation standardisée de processus d'évaluation d'audience

Sommaire

4.1	Introduction	58
4.2	Définitions	58
4.2.1	Évaluation d'audience	58
4.2.2	Système d'évaluation d'audience	60
4.3	Besoin d'un langage commun pour la description des processus	60
4.4	BPMN, une notation standard pour la description formelle de processus	61
4.4.1	Pourquoi BPMN ?	61
4.4.2	Éléments de base de BPMN	62
4.5	Modèles de processus caractérisant un cours en face-à-face	65
4.5.1	Modèle de processus d'un cours magistral	66
4.5.2	Modèle de processus d'un cours interactif	66
4.6	Modèles de processus d'évaluation d'audience supportés par les SVI	67
4.6.1	Le modèle de processus de la situation nominale	69
4.6.2	Le modèle de processus de l'approche "Instruction par les pairs"	69
4.7	Le processus d'évaluation d'audience en N phases	72

4.1 Introduction

Nous introduisons dans ce chapitre le concept d'évaluation d'audience ainsi qu'un nouveau processus relevant de ce concept et supportant l'ensemble des d'activités et des stratégies caractéristiques de l'évaluation formative. Ces nouvelles notions, premières contributions issues de nos travaux, visent la fourniture d'un socle permettant l'implantation d'un système conforme aux spécifications du système cible introduit dans la section 3.5.

4.2 Définitions

Cette section a pour objectif la définition claire et synthétique de notre système cible dont les caractéristiques ont été décrites dans le chapitre précédent.

4.2.1 Évaluation d'audience

Pour définir leur cadre théorique de l'évaluation formative, Black et Wiliam (2009) identifient les stratégies relevant de l'évaluation formative et intègrent ces stratégies dans trois processus clés de l'apprentissage (cf. section 2.6.3). L'évaluation formative est modélisée sous forme de stratégies s'intégrant dans un ensemble de processus d'apprentissage.

La revue de littérature de Roschelle *et al.* (2004) portant sur les systèmes de votes interactifs (SVI) propose en synthèse un modèle générique décrivant comment l'agrégation de technologies dans la classe permet d'atteindre des résultats profitables en terme d'apprentissage. Le modèle CATAALYST (*Classroom Aggregation Technology for Activating and Assessing Learning and Your Students' Thinking*) synthétisé en Figure 3 présente un processus générique décrivant comment les actions de l'enseignant combinées avec une technologie dotée de quelques fonctionnalités clés affectent les étudiants dans une classe. Le modèle CATAALYST ne préjuge pas des types d'interactions mis en œuvre par le dispositif technologique utilisé ni du type de feedback restitué aux étudiants.

La notion de processus apparaît donc centrale pour définir tout ce qui relève de l'évaluation formative, sur le plan théorique comme sur le plan de sa mise en œuvre à

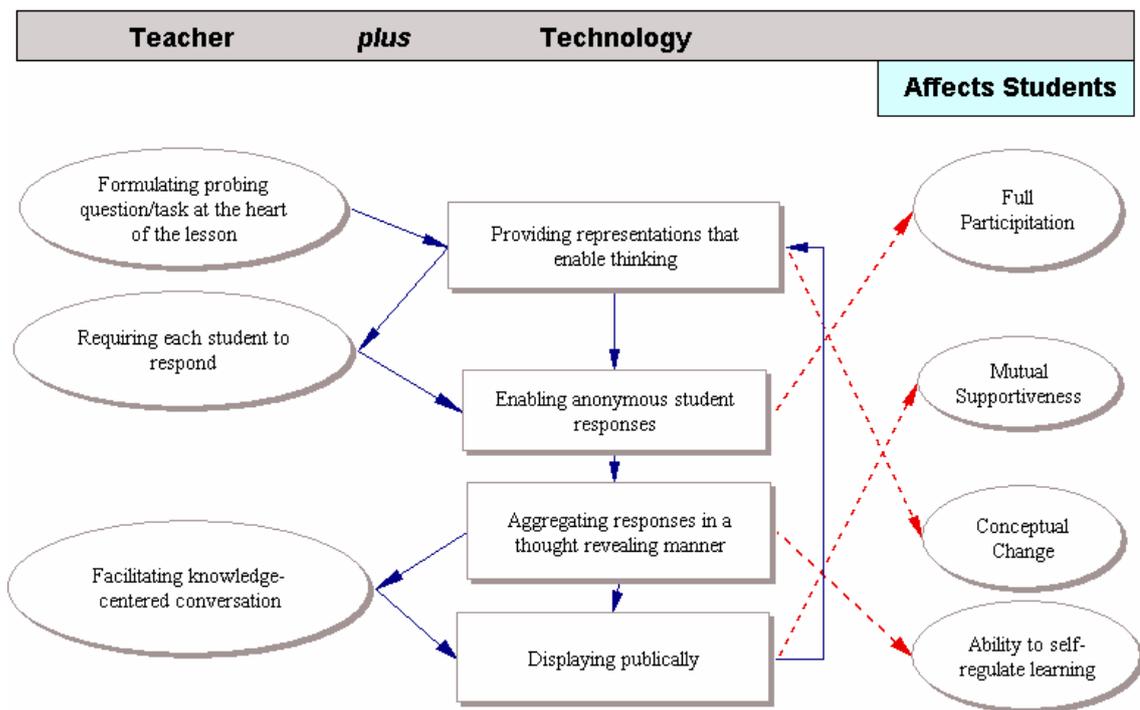


Figure 3. Le modèle CATAALYST de Roschelle *et al.* (2004)

travers des dispositifs technologiques comme les SVI. Nous proposons alors de définir le concept d'évaluation d'audience dans le contexte de l'enseignement supérieur de la manière suivante :

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.

Cette définition place la notion de processus au cœur du concept. La première partie de la définition associe la notion d'évaluation d'audience à un processus d'évaluation formative, c'est-à-dire à un ensemble d'activités s'inscrivant dans les stratégies caractéristiques de l'évaluation formative. La deuxième partie de la définition précise le contexte dans lequel ce processus est appliqué : un cours en face-à-face délivré à un grand nombre d'étudiants réunis par exemple dans un amphithéâtre d'université. Enfin, la troisième partie de la définition précise que le processus s'adresse simultanément à l'ensemble des étudiants. Cette précision permet d'exclure tout ensemble d'activités qui ne solliciterait qu'une partie des étudiants comme, par exemple, l'évaluation pendant le cours d'un groupe de quelques étudiants

présentant un projet ou un exposé. Cette définition permet de clairement affirmer que les évaluations formatives mises en œuvre à l'aide des systèmes de votes interactifs sont des évaluations d'audience.

4.2.2 Système d'évaluation d'audience

À partir de la définition du concept d'évaluation d'audience, nous introduisons ici la définition d'un système supportant la mise en œuvre de ce concept.

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

Les SVI sont des SEA ; s'il était possible de définir un modèle de maturité des SEA, les SVI seraient les SEA de niveau 1.

La première partie de la définition établit le lien direct entre SEA et processus d'évaluation d'audience. La deuxième partie de la définition permet d'intégrer les SVI dans l'ensemble des SEA en signifiant que tout SEA est au minimum un SVI.

La définition du concept étant posée, nous proposons de contribuer à sa clarification en identifiant et en décrivant les processus d'évaluation d'audience s'appuyant sur les SVI. Cette démarche pose la question de la manière de décrire de tels processus.

4.3 Besoin d'un langage commun pour la description des processus

Les processus décrits par Black et Wiliam (2009) ou Roschelle *et al.* (2004) sont représentés sous forme tabulaire (voir Tableau 1) ou graphique (voir Figure 3) présentant de nombreux points communs : ils utilisent une colonne par acteur impliqué dans le processus et les activités apparaissent comme des items de chaque colonne. Ces partis pris rappellent fortement les notations standards telles que BPMN¹ ou les diagrammes d'activité UML². Pour autant, ces présentations formelles ne s'appuient sur aucune de ces notations standards et ne facilitent donc pas le partage et la compréhension des différents processus formalisés par l'ensemble des parties prenantes (théoriciens, praticiens, informaticiens, etc.). Nous proposons

1. Business Process Model and Notation
2. Unified Modeling Language

l'utilisation d'un langage standard pour la description de processus afin de faciliter cette communication entre les différents acteurs ayant à échanger sur les processus relatifs à l'évaluation formative.

4.4 BPMN, une notation standard pour la description formelle de processus

4.4.1 Pourquoi BPMN ?

Roiguet *et al.* (2010) déclarent que l'une des principales barrières déconnectant les sciences de l'éducation des sciences informatiques provient de la difficulté pour les enseignants et éducateurs de décrire les activités d'apprentissages en processus et *workflows* correspondant au niveau auquel les informaticiens sont capables de les traiter.

Il existe cependant de nombreux langages pour la description d'éléments liés à des situations d'apprentissage (ontologies pour les domaines de connaissances, langages de méta-données pour les ressources pédagogiques, etc.) (Laforcade, 2004). Certains de ces langages, les *Educational Modeling Languages* (EML), sont dédiés à la modélisation d'activités pédagogiques. Plus précisément, un EML permet de décrire le contenu et le processus d'une unité d'apprentissage d'un point de vue pédagogique dans l'objectif de permettre la réutilisation et l'interopérabilité (Rawlings *et al.*, 2002).

IMS-LD (IMS, 2003b), le langage normalisé par l'IMS³ à la suite des travaux réalisés par l'*Open University* des Pays-Bas sur les EML, est un langage visant le support de la diversité et de l'innovation pédagogique en promouvant les échanges et l'interopérabilité des matériaux du *e-learning*. IMS-LD est un langage XML ; il est donc difficilement utilisable directement par les enseignants. Le guide des bonnes pratiques accompagnant la spécification d'IMS-LD (IMS, 2003a) propose l'utilisation d'un langage visuel pour décrire les situations pédagogiques en phase initiale ; le guide utilise dans ses exemples les diagrammes d'activités du langage UML (OMG, 2000). Différents travaux se sont intéressés à l'utilisation de langages visuels pour la description de situations d'apprentissage (Laforcade, 2004; Nodenot, 2007) en vue d'une traduction éventuelle en IMS-LD : Paquette *et al.* (2006) et Martinez-Ortiz

3. <http://www.imsglobal.org/>

et al. (2009) proposent des langages ad hoc, alors que Dodero *et al.* (2008) préfèrent s'appuyer sur le langage UML et proposent des techniques de transformation d'UML à IMS-LD.

Le langage BPMN 2.0 (OMG, 2011) est un langage visuel de description de modèles de processus métier. Il est normalisé, comme UML, par l'OMG qui a souhaité, avec BPMN, atteindre deux objectifs principaux :

- obtenir un langage le plus lisible possible pour l'ensemble des parties prenantes (les personnes du métier, les informaticiens, etc.) ;
- fournir un langage standard de visualisation pour des processus métiers définis à l'aide d'un langage exécutable tel que WSBPEL (OASIS, 2007).

L'OMG (2011) indique que BPMN 2.0 regroupe l'ensemble des bonnes pratiques ayant émergé de l'utilisation de langages antérieurs (diagrammes d'activités UML, UML EDOC Business Processes, IDEF, ebXML BPSS, diagrammes *Activity-Decision Flow*, RosettaNet, etc.). D'autre part, BPMN commence à susciter la curiosité de chercheurs s'intéressant aux environnements informatiques pour l'apprentissage humain (EIAH) : Da Costa et Schneider (2015) invitent à l'expérimentation du langage pour décrire des processus correspondant à des situations d'apprentissage afin de mettre BPMN à l'épreuve dans ce champs de recherche.

Les ambitions affichées de l'OMG sur BPMN 2.0 et la démarche proposée par Da Costa et Schneider (2015) ont orienté notre choix de BPMN 2.0 pour décrire les processus relatifs à l'évaluation d'audience. La section suivante introduit les éléments de base du langage en se limitant aux notions BPMN utilisées dans le cadre de nos travaux.

4.4.2 Éléments de base de BPMN

Dans cette section, les noms communs écrits avec leur première lettre en capitale correspondent à des mots clés du langage BPMN.

BPMN permet de modéliser des Processus métier de bout-en-bout. Parmi les sous-modèles qui permettent de décrire un modèle bout-en-bout, BPMN introduit les Processus et les Collaborations.

BPMN distingue deux types de Processus : les Processus privés et les Processus publics. Les Processus privés sont internes à une organisation. Ils sont souvent désignés dans la littérature relative au *Business Process Management* (BPM) sous

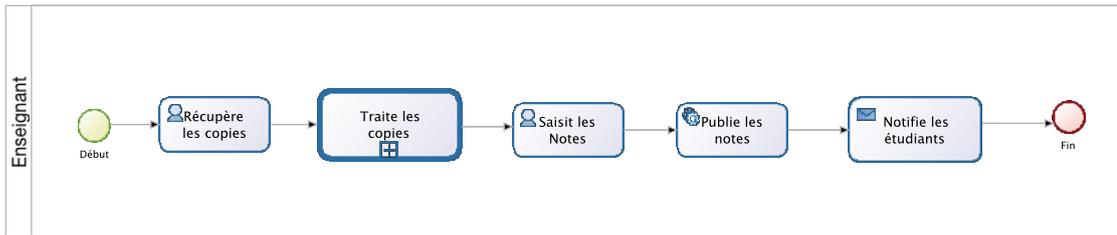


Figure 4. Exemple de Processus privé

le terme "workflow" ou "Processus BPM". Dans le domaine des Web Services, un Processus privé correspond à une Orchestration de services. Un Processus privé est éventuellement exécutable s'il a été conçu en respectant les exigences rendant possible l'exécution. La Figure 4 représente un exemple de Processus privé concernant la gestion des copies et des notes d'un enseignant du secondaire suite à une épreuve d'évaluation sur feuille adressée à ses élèves.

Un Processus public représente les interactions entre un Processus privé et un autre Processus ou Participant. Seules les activités qui sont utilisées pour communiquer avec le ou les autres Participants sont incluses dans le Processus public. Celui-ci fait donc apparaître les Flux de Message et l'ordre de ces Flux nécessaires à la communication avec le Processus externe. Les Processus publics peuvent être modélisés de manière indépendante ou dans le contexte d'un modèle de Collaboration afin de montrer les Messages entre le Processus public et les autres Participants.

Une Collaboration décrit les interactions entre au moins deux entités métier. Une Collaboration contient habituellement deux Piscines représentant les Participants dans la Collaboration. Les Messages échangés entre les Participants sont représentés à l'aide de Flux de Messages qui connectent deux Piscines. Une Collaboration peut être une composition de deux Processus publics communiquant entre eux. La Figure 5 représente une Collaboration entre notre enseignant du secondaire et ses élèves dans le cadre du traitement de ses copies.

Le tableau 7 contient une liste des éléments de modélisation de base utilisés pour la description des processus.

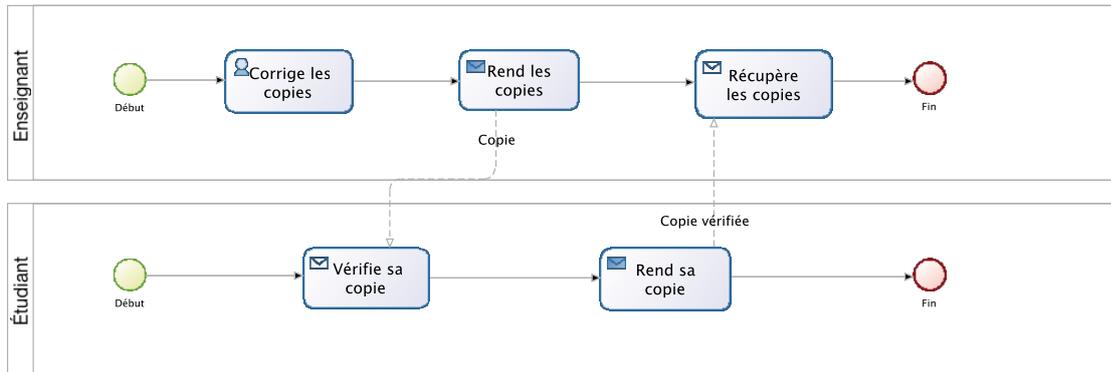
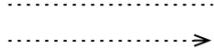
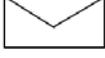


Figure 5. Exemple de Collaboration

Table 7. Éléments de base de BPMN 2.0 (OMG, 2011)

Élément	Description	Notation
Évènement	Un Évènement est quelque chose qui se produit au cours d'un Processus. Les Évènements affectent le flux du modèle et ont généralement une cause (déclencheur) ou un impact (un résultat). Les Évènements sont représentés par des cercles dans lesquels il est possible d'insérer des marqueurs pour distinguer les différents types de déclencheurs ou résultats. Il y a trois types d'Évènements dépendant du moment où se produit l'évènement dans le Processus : Évènements de Début, Intermédiaires et de Fin.	
Activité	Une Activité est un terme générique pour décrire un travail qu'un Participant réalise au cours d'un Processus. Une Activité peut être atomique (Tâche) ou composée (Sous Processus). Les Activités sont représentées à l'aide de rectangles aux bords arrondis.	
Porte	Une Porte est utilisée pour contrôler la divergence et la convergence des Flux de Séquence dans un Processus. Les Portes gèrent ainsi les branchements, les embranchements, les fusions et les jointures de chemins. Les marqueurs insérés dans les losanges représentant les Portes permettent d'indiquer le type de contrôle de comportement.	

Table 7. Éléments de base de BPMN 2.0 (OMG, 2011)

Élément	Description	Notation
Flux de séquence	Un Flux de séquence est utilisé pour indiquer l'ordre dans lequel les Activités sont exécutées dans un Processus.	
Flux de message	Un Flux de Message est utilisé pour indiquer un flux de messages entre deux Participants. Dans BPMN, deux Participants sont représentés par deux Piscines distinctes.	
Association	Une Association est utilisée pour lier une information à une Annotation ou à un autre Artefact BPMN. Une Association peut être fléchée pour indiquer le sens de l'Association si nécessaire.	
Piscine	Une Piscine est la représentation graphique d'un Participant dans une Collaboration. Une piscine peut afficher les détails internes du Processus qu'elle décrit ou ne rien afficher du tout et se présenter alors comme une boîte noire.	
Message	Un Message est utilisé pour décrire le contenu d'une communication entre deux Participants.	
Groupe	Un Groupe est le regroupement d'éléments graphiques dans une même Catégorie. Un Groupe n'affecte pas les Flux de Séquence à l'intérieur du Groupe. Le nom de la Catégorie apparaît sur le diagramme comme le label d'un groupe.	
Annotation (liée à l'aide d'une association)	Les Annotations sont un moyen de fournir de l'information textuelle complémentaire pour le lecteur d'une diagramme BPMN.	

Nous compléterons les explications relatives à la notation BPMN en cas de besoin dans la suite du document.

4.5 Modèles de processus caractérisant un cours en face-à-face

Nous proposons dans cette section des modèles correspondant aux processus mis en jeux dans le contexte de cours en face-à-face. Les modèles couvrent différentes

situations :

- le cours en face-à-face dispensé sans interaction ; nous désignerons ce type de cours par l'expression "cours magistral" ;
- le cours en face-à-face dispensé avec alternance de présentations et d'évaluations ; nous désignerons de type de cours par l'expression "cours interactif".

4.5.1 Modèle de processus d'un cours magistral

La Figure 6 décrit le processus relatif à un cours magistral (CM). L'enseignant et les étudiants ne sont liés que par les deux événements de début et de fin du cours. L'enseignant présente les contenus de son cours tandis que les étudiants prennent des notes. L'évènement de début de l'enseignant est représenté par le symbole  signifiant que l'évènement est déclenché par un facteur temps. Dans notre cas, c'est la date et l'heure du début du cours qui marque le début du processus. Cependant, le cours ne commence que quand l'enseignant notifie explicitement aux étudiants que le cours commence. Une activité accompagnée du symbole  est une tâche de notification. Une tâche de notification permet d'envoyer un message à un autre participant de la collaboration. Dans notre modèle, le message est la notification adressée aux étudiants. La réception de ce message par les étudiants marque le début du processus côté étudiant. L'évènement de début d'un processus déclenché par un message est représenté par le symbole . Les étudiants reçoivent à la fin du cours une notification de fin de cours marquant le passage à la fin du processus. Un évènement intermédiaire correspondant à la réception d'un message est représenté par le symbole .

Enfin notons sur ce diagramme les tâches réalisées par les personnes : ces activités sont marquées du symbole  signifiant que l'activité est une tâche exécutée par un acteur humain, celui-ci pouvant être aidé d'un logiciel pour exécuter sa tâche.

4.5.2 Modèle de processus d'un cours interactif

La Figure 7 décrit le processus relatif à un cours interactif. Le cours interactif est représenté comme une succession de séquences.

Un enseignant commence une séquence qui est soit une présentation du contenu du cours, soit une évaluation de l'audience. Le symbole  représente une porte de type "Ou exclusif" et permet d'exprimer le choix exclusif de l'enseignant portant sur

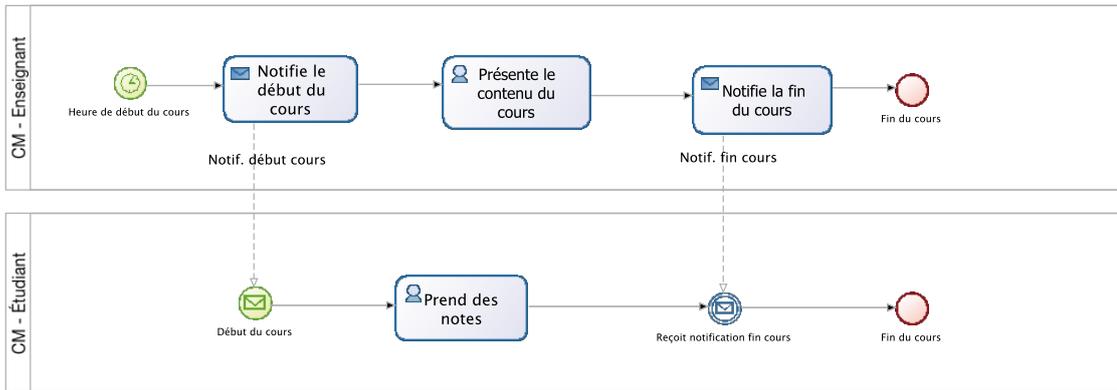


Figure 6. Modèle du processus Cours Magistral

le type de la séquence en cours. Une autre porte XOR est utilisée dans le processus de l'enseignant permettant d'orienter le processus vers la fin du cours ou vers le début d'une nouvelle séquence. L'activité "Évalue l'audience" est représentée sous une forme particulière : le contour est plus épais que dans le cas d'une activité standard et sur la base du cadre apparaît le symbole . Une activité sous cette forme représente un sous-processus ayant vocation à être détaillé sur un diagramme dédié. Cette approche permet de modéliser à différents niveaux de granularité. Dans notre cas, les sous-processus d'évaluation d'audience côté enseignant et côté étudiant seront modélisés dans la section suivante.

Les étudiants, de manière symétrique, alternent les séquences où ils participent à l'évaluation de l'audience avec les séquences de présentation du contenu du cours pendant lesquelles ils prennent des notes.

Notons que le modèle proposé est suffisamment générique pour représenter aussi bien une approche guidée par les évaluations (approche du type "TEFA") qu'une approche guidée par la présentation des contenus (approche "Instruction par les pairs").

4.6 Modèles de processus d'évaluation d'audience supportés par les SVI

Nous avons décrit dans la section 3.3.2 les différentes approches de mise en œuvre d'évaluation formative à l'aide des systèmes de votes interactifs (SVI). Nous disposons, avec ces descriptions, de tous les éléments caractérisant des processus

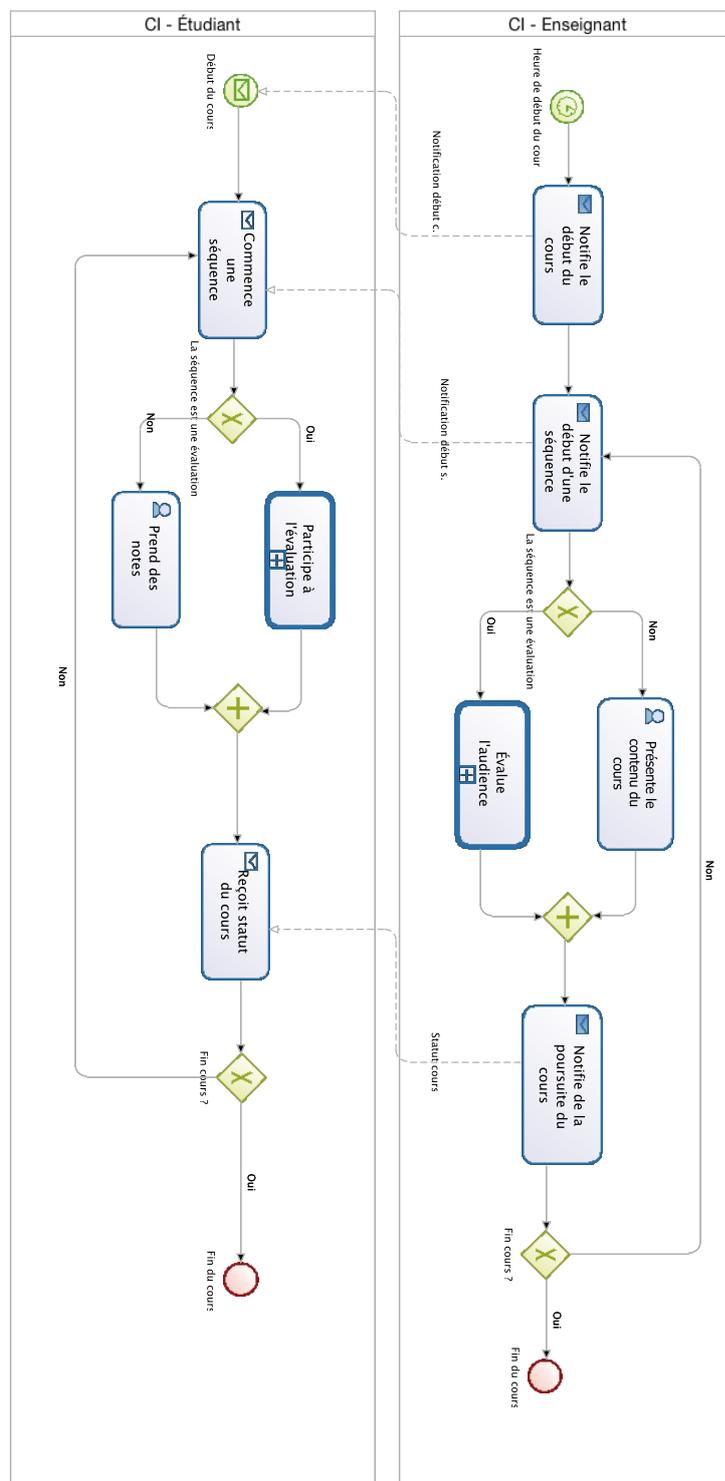


Figure 7. Modèle du processus Cours Interactif

relevant directement de ce que nous avons défini comme de l'évaluation d'audience. En modélisant les processus d'évaluation formative supportés par les SVI, nous modélisons les processus supportés par les systèmes d'évaluation d'audience les plus courants à ce jour. Nous présentons dans cette section les modèles des deux processus les plus sollicités pour la mise en œuvre des SVI : le processus correspondant à ce que nous avons baptisé "Situation nominale" et celui correspondant à l'approche "Instruction par les pairs".

4.6.1 Le modèle de processus de la situation nominale

Nous utiliserons dans la suite du document l'expression "évaluation d'audience nominale" pour désigner la situation nominale de mise en œuvre des SVI. La Figure 8 présente le modèle de processus d'évaluation d'audience nominale.

Le processus est initié par l'enseignant lorsqu'il démarre une nouvelle séquence d'évaluation. Il pose alors une question aux étudiants qui se traduit sur le modèle par l'envoi d'un message contenant la question. Ce message déclenche le processus côté étudiant. Chaque étudiant répond à la question et transmet sa réponse au processus enseignant. La collecte des réponses est une tâche de réception de message particulière : elle se répète tant que des réponses sont à réceptionner. Le caractère itératif de la tâche est représenté par le symbole . C'est côté enseignant que le traitement des réponses est effectué après réception de l'ensemble des réponses. Le symbole  sur l'activité de traitement indique que la tâche est automatisée, ce qui est bien le cas dans les situations faisant usage des SVI. Le feedback obtenu à l'issue du traitement des réponses est présenté aux étudiants via une tâche de notification. Après réception du feedback par les étudiants, une discussion est lancée. La discussion est modélisée sous la forme d'une tâche d'animation côté enseignant et sous la forme de deux tâches pouvant s'exécuter en parallèle côté étudiant : la participation à la discussion et la prise de notes.

4.6.2 Le modèle de processus de l'approche "Instruction par les pairs"

Nous utiliserons dans la suite du document l'expression "évaluation d'audience Instruction par les pairs" pour désigner la situation de mise en œuvre des SVI suivant l'approche "Instruction par les pairs". La Figure 9 présente le modèle du processus

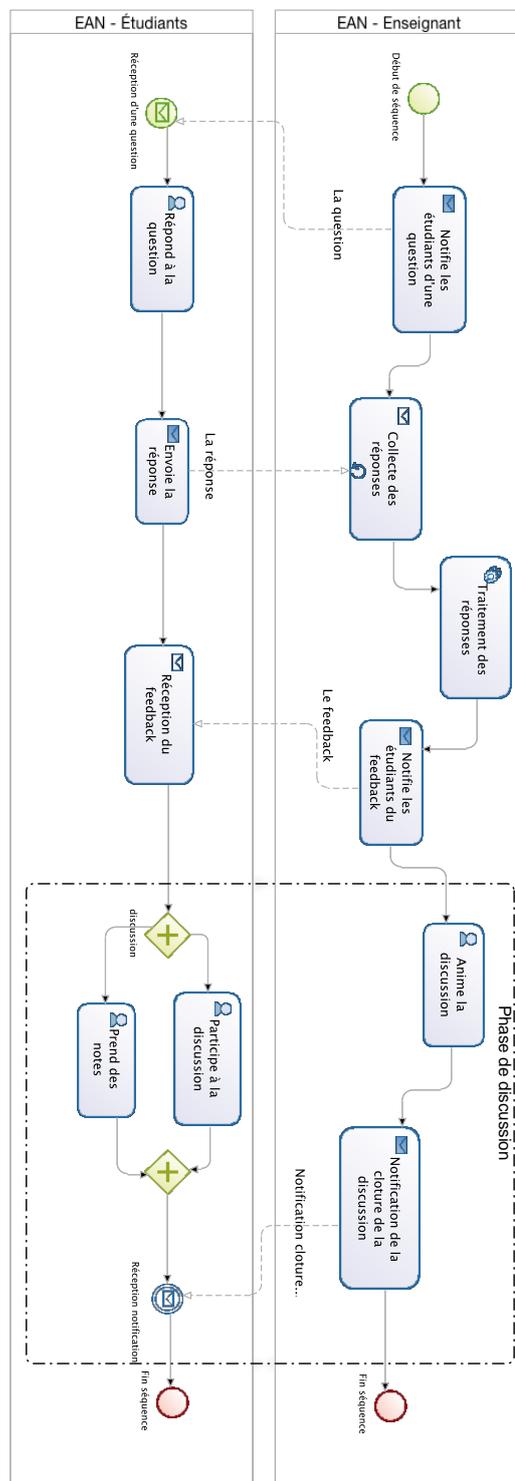


Figure 8. Modèle du processus d'évaluation d'audience nominale

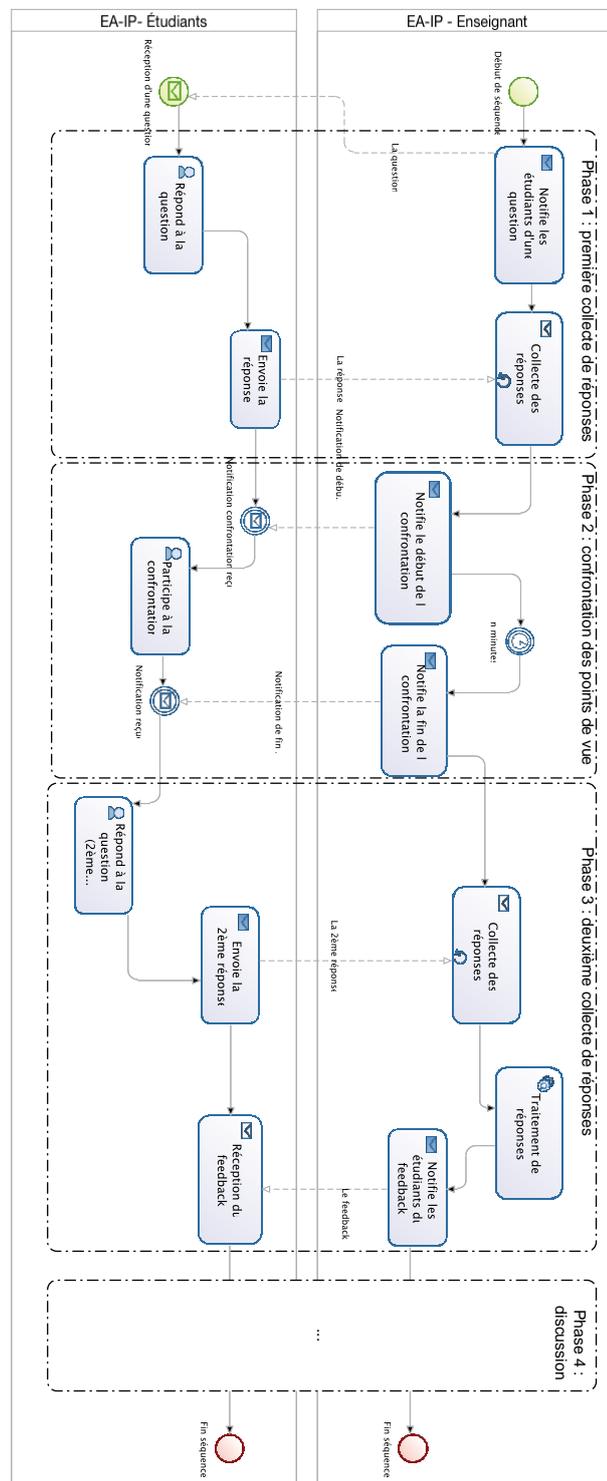


Figure 9. Modèle du processus d'évaluation d'audience "Instruction par les pairs"

d'évaluation d'audience Instruction par les pairs.

Le processus est modélisé suivant quatre phases représentées par des groupes BPMN. La première phase correspond à la pose de la question et à la première collecte de réponses. La deuxième phase permet aux étudiants de confronter leur point de vue. La troisième phase correspond à la deuxième collecte de réponses donnant lieu à un feedback fourni aux étudiants. La dernière phase correspond à la phase de discussion clôturant la séquence. Cette dernière phase n'est pas explicitée sur le diagramme car elle correspond exactement à la phase de discussion détaillée sur processus d'évaluation d'audience nominale (cf. Figure 8).

4.7 Le processus d'évaluation d'audience en N phases

Les processus supportés par les SVI décrits précédemment sont limités en terme de mise en œuvre d'évaluations formatives. En effet, l'évaluation par les pairs (activité de référence A-4) n'apparait dans aucun des processus supportés par les SVI et ceux-ci ne proposent ni la saisie de réponses sous forme de texte libre (I-ST), ni la fourniture de feedbacks textuels (I-TF) en raison des limites des SVI actuels (cf. section 3.3.3).

Afin de dépasser ces limites, nous avons conçu un nouveau processus d'évaluation d'audience baptisé "processus en N phases" qui consiste à découper la soumission d'une réponse par un étudiant en plusieurs phases. La Figure 10 fournit une description BPMN du processus en N phases dans sa mise en œuvre nominale comprenant quatre phases.

La première phase correspond à la pose de la question par l'enseignant et au recueil des réponses fournies par les étudiants. Le processus est conçu de telle sorte que les réponses puissent contenir des éléments saisis sous forme de texte libre. Par abus de langage, nous désignerons dans la suite du manuscrit par l'expression "réponse textuelle" toute réponse contenant au moins un élément fourni sous forme de texte libre.

La deuxième phase a été conçue pour intégrer explicitement la confrontation de points de vue au cœur du système cible : à chaque étudiant est proposée une réponse différente de la sienne et fournie par un autre étudiant. Cette phase propose un premier feedback textuel à l'étudiant à travers la proposition d'une réponse

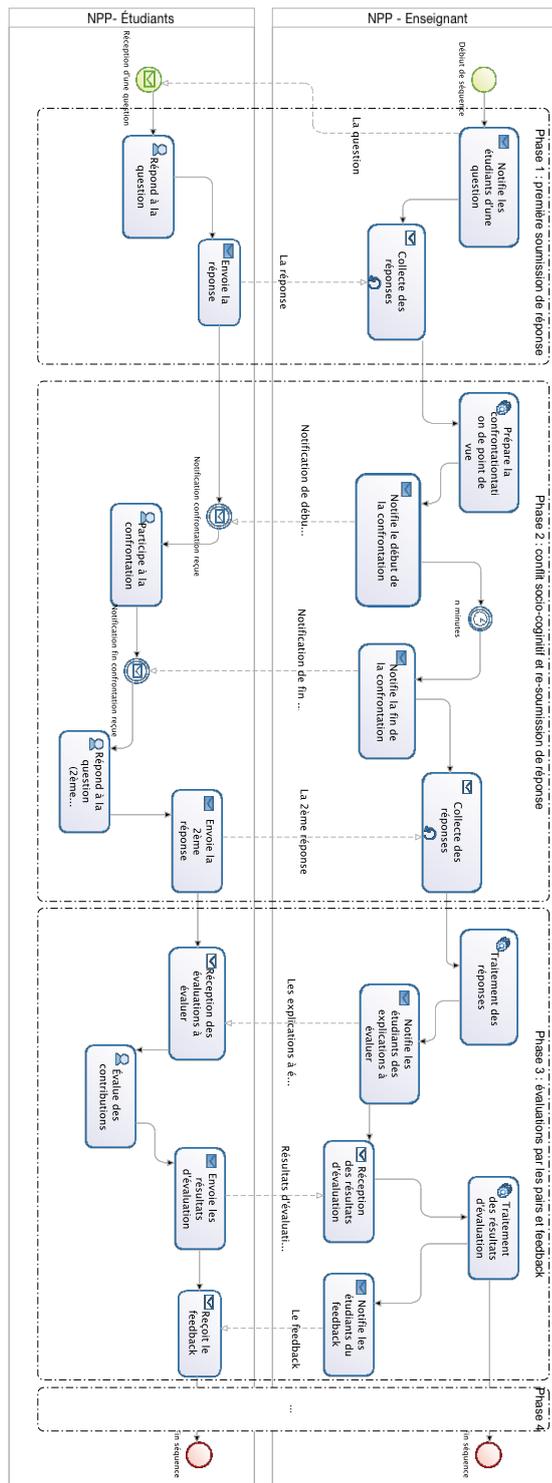


Figure 10. Modèle du processus de soumission de réponses en N phases

alternative. L'étudiant est alors mis en situation de conflit socio-cognitif : il doit confronter sa réponse à la réponse différente proposée par l'un de ses pairs. L'étudiant peut alors conforter ou modifier sa réponse. Les réponses alternatives sont proposées de manière anonyme afin de neutraliser tout effet néfaste d'influence relationnelle.

La troisième phase est dédiée à l'évaluation des réponses. L'activité d'évaluation des réponses textuelles est déléguée aux étudiants. Chaque étudiant reçoit un lot de réponses à évaluer ; les résultats des évaluations fournies par les étudiants sont traités afin de fournir un feedback. Cette approche règle d'une part la question de l'évaluation en temps réel de réponses textuelles en parallélisant ce traitement auprès des différents étudiants, et d'autre part intègre de fait l'activité d'évaluation par les pairs.

La quatrième phase est la phase de discussion animée par l'enseignant suite à la présentation du feedback aux étudiants. Ces derniers participent à la discussion en prenant éventuellement des notes. Cette dernière phase n'est pas explicitée sur le diagramme car elle correspond exactement à la phase de discussion détaillée sur processus d'évaluation d'audience nominale (cf. Figure 8).

Une variante de la mise en œuvre nominale est la répétition de la seconde phase pour confronter chaque étudiant à différentes réponses alternatives. Dans ce cas, le processus devient un processus en N phases où N dépend du nombre d'itérations sur la phase 2.

4.8 Conclusion

Ce chapitre introduit la notion d'évaluation d'audience et fournit ainsi un cadre précis pour nos travaux : le système cible décrit dans la section 3.5 est un système d'évaluation d'audience dépassant les limites des SVI. Les processus caractérisant l'évaluation d'audience supportée par les SVI ont été décrits en utilisant le langage de modélisation de processus BPMN 2.0. L'utilisation de ce langage standard offre de nouvelles perspectives d'échanges entre les différents acteurs de l'évaluation formative, qui peuvent à présent discuter et comparer des processus sur la base d'un langage commun. Enfin, le processus en N phases, conçu de manière à enrichir l'expérience d'évaluation d'audience par rapport aux expériences issues de l'utilisation des SVI, a été présenté en détail dans la section précédente.

L'implantation du processus en N phases dans sa forme nominale a été réalisée au sein du système Tsaap-Notes au terme d'un cycle de conception et de développement

composé de trois itérations. Le chapitre suivant décrit de manière détaillée les résultats obtenus à l'issue de chaque itération. Chacun de ces résultats a fait l'objet d'une expérimentation ; les expérimentations et l'analyse de leurs résultats sont présentées dans les chapitres 6 et 7.

Chapitre 5

Conception et mise en œuvre du SEA Tsaap-Notes

Sommaire

5.1	Introduction	78
5.2	Choix d'architecture technique	78
5.2.1	Exigences non fonctionnelles	78
5.2.2	Une architecture bâtie sur Java EE	80
5.2.3	Les patrons de conception mobilisés	82
5.2.4	Le <i>framework</i> de développement	85
5.2.5	Focus sur les interfaces clientes	87
5.2.6	Le système de gestion de bases de données relationnelles	87
5.3	Itération 1 - SVI, micro-blogging et prise de notes collaborative : Tsaap-Notes	87
5.3.1	Introduction	87
5.3.2	Impacts sur les processus	88
5.3.3	Aperçu de Tsaap-Notes	88
5.3.4	Lier les ressources du cours avec les notes	90
5.3.5	Les vues personnalisées réutilisables	92
5.3.6	Le système de votes interactifs	93
5.3.7	Synthèse des résultats	94
5.4	Itération 2 - La génération semi-automatique de tests de révisions	95
5.4.1	Introduction	95

5.4.2	Tsaap-Notes et la génération de tests d'auto-évaluation . . .	97
5.4.3	Synthèse des résultats	98
5.5	Itération 3 - Le SEA Tsaap	100
5.5.1	Introduction	100
5.5.2	La mise en œuvre du processus en N phases	101
5.5.3	Focus sur l'algorithme d'association de réponses pour le conflit socio-cognitif	102
5.5.4	La génération semi-automatique de tests améliorée	106
5.5.5	Synthèse des résultats	107
5.6	Conclusion	107

5.1 Introduction

Ce chapitre décrit la conception et la mise en place du système d'évaluation d'audience (SEA) Tsaap-Notes qui implante les concepts et processus définis dans le chapitre précédent afin de lever les verrous identifiés dans la section 3.6. La présentation du système Tsaap-Notes est répartie en quatre sections. La première section présente les choix d'architecture technique. Trois sections sont ensuite consacrées respectivement aux trois itérations majeures de la conception et de la mise en œuvre de Tsaap-Notes.

5.2 Choix d'architecture technique

5.2.1 Exigences non fonctionnelles

Les choix d'architecture technique ont été faits pour respecter différentes exigences décrites ci-dessous.

5.2.1.1 L'utilisation du réseau Internet

Nous souhaitons que le système mis en œuvre soit facile à déployer dans les amphithéâtres. Les enseignants ne doivent pas transporter de matériel complémentaire et ne doivent pas créer de réseau ad hoc pour l'utilisation du système. Nous faisons donc l'hypothèse que les établissements supportant le déploiement des SEA mettent à disposition dans les amphithéâtres un réseau WIFI

fournissant un accès à Internet haut débit et fiable. Ces hypothèses nous permettent d'exprimer une première exigence.

Le système mis en œuvre s'appuie sur le réseau Internet pour assurer la communication entre ses différentes couches.

5.2.1.2 Le support des dispositifs connectés à Internet

Le système doit permettre aux étudiants et à l'enseignant d'utiliser tout type de dispositif capable de se connecter à Internet et doté d'un clavier alphanumérique : ordinateur portable, tablette ou smartphone.

Cette exigence est notamment compatible avec une approche "Apportez votre Équipement personnel de Communication¹" (AVEC) autorisant les utilisateurs à participer aux activités supportées par le système à l'aide de leur propre matériel.

Cette approche lève cependant la problématique des utilisateurs n'ayant pas de matériel connecté à Internet dans les amphithéâtres, soit parce qu'ils ne possèdent pas un tel matériel, soit parce qu'ils ne souhaitent pas le transporter et l'utiliser en cours. Nous considérons que cette problématique n'est pas dans le périmètre de la thèse. Il appartient à chaque établissement de régler cette question. Nous aborderons la manière dont l'établissement d'expérimentation a réglé le problème dans le chapitre 6.

5.2.1.3 Le client Web

Afin de supporter un maximum de dispositifs connectés à Internet, le choix du navigateur comme client universel s'est imposé.

Les utilisateurs accèdent au système à l'aide d'un navigateur Web. Les navigateurs Web supportés sont les navigateurs supportés par les dispositifs mobiles les plus courants : Safari et Chrome dans leurs versions récentes.

En synthèse des trois premières exigences formulées, le système que nous avons conçu est une application Web en accord avec son temps.

1. En anglais : *Bring Your Own Device*, soit BYOD en abrégé

5.2.1.4 Montée en charge

Le système doit permettre aux étudiants d'un amphithéâtre de répondre de manière quasi-simultanée à une question posée par l'enseignant. Le système doit tenir la charge correspondant à une telle situation y compris dans des cas limites où l'amphithéâtre contient plusieurs centaines d'étudiants.

Le système doit supporter la charge correspondant à plusieurs centaines d'étudiants connectés en même temps.

5.2.1.5 Sécurité et protection des données personnelles

Le système que nous avons conçu stocke des informations personnelles des étudiants et des enseignants.

Le système doit être sécurisé et doit répondre aux exigences de la CNIL en terme de traitement des données personnelles des utilisateurs.

5.2.1.6 Un outillage Open Source

Les bibliothèques tierces et autres *frameworks* utilisés pour le développement du système doivent être Open Source et distribués sous une licence de type "no copyleft" ou "weak copyleft" afin qu'aucune contrainte sur le code du système ne soit imposée par l'utilisation d'un outil tiers. La plupart des bibliothèques tierces utilisées sont des bibliothèques distribuées sous licence Apache.

5.2.2 Une architecture bâtie sur Java EE

Le système à bâtir est une application Web devant répondre à des exigences de sécurité et de robustesse. L'environnement Java Entreprise Edition (Java EE) a été conçu pour la mise en œuvre de telles applications.

Java EE est un ensemble de spécifications destinées à faciliter le développement d'applications distribuées répondant aux exigences des applications d'entreprises. Les spécifications Java EE décrivent la plateforme Java EE représentée sur la Figure 11 et permettent ainsi à différents éditeurs de fournir les implémentations des différents composants d'une application compatible Java EE.

La plateforme Java EE décrit deux grandes familles de conteneurs : les conteneurs clients et les conteneurs serveurs. Le protocole standard de communication entre les composants Java EE étant HTTP(S), le client d'une application Java EE peut

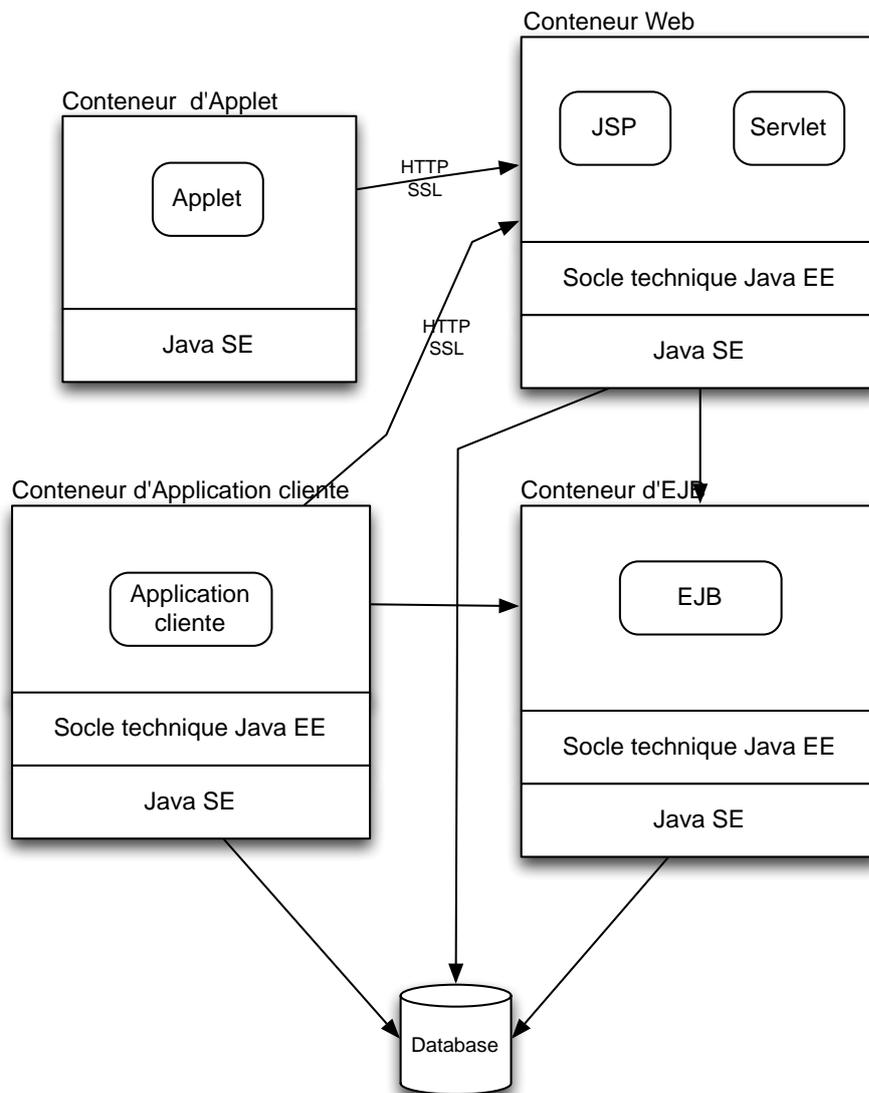


Figure 11. La plateforme Java EE

être un simple client Web. De fait, les conteneurs clients (conteneur d'Applet, conteneur d'applications clientes) ne sont pas indispensables au fonctionnement d'une application Web Java EE.

Un serveur d'application Java EE est une implémentation d'un ou des deux conteneurs serveurs Java EE. Tomcat² est une implémentation très populaire du conteneur Web, alors que Glassfish³ est l'implémentation de référence des deux conteneurs (Web et EJB). Plusieurs instances de serveurs d'applications Java EE peuvent être déployées pour supporter une charge grandissante.

Une application serveur Java EE se déploie donc dans un serveur d'applications Java EE. Elle peut être aisément sécurisée en permettant d'implémenter tout type de politique de sécurité exigée par la CNIL (service d'authentification sécurisé, URL sécurisées, encryptage des données sensibles, etc.).

Java EE 6 introduit le concept de profil d'applications pour qualifier les applications en fonction du type de conteneur nécessaire à leur déploiement. Le conteneur d'EJB, dédié à la gestion d'objets métiers distribués, n'est pas indispensable au fonctionnement d'une application serveur Java EE. Le profil Web défini dans Java EE 6 décrit les applications serveurs déployables dans un simple conteneur Web. La logique métier est dans ce cas décrite dans des objets gérés par la JVM exécutant le processus du conteneur. Nous utiliserons l'expression "application web" pour désigner une application serveur Java EE ayant le profil Web.

Le système que nous avons développé est une application Web respectant quelques patrons de conceptions essentiels décrits dans la section suivante.

5.2.3 Les patrons de conception mobilisés

5.2.3.1 Modèle Vue Contrôleur revisité

Le patron Modèle Vue Contrôleur (MVC) préconise le découpage d'une application en trois niveaux de responsabilité afin de faciliter l'évolution et la maintenance de l'application. La Figure 12 offre une représentation schématique de MVC.

Le niveau de responsabilité "Modèle" est en charge de l'expression de la logique métier. Dans une application Web, le modèle est implanté sous la forme d'objets métiers. Le niveau de responsabilité "Vue" est en charge de tous les aspects interfaces.

2. <http://tomcat.apache.org/>

3. <https://glassfish.java.net/>

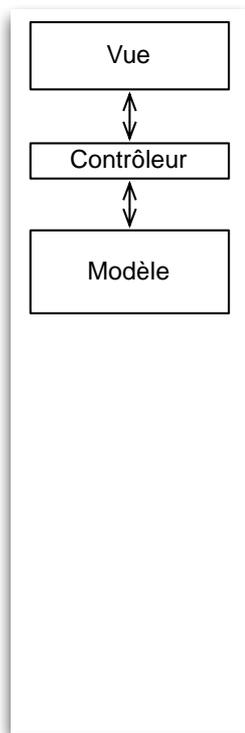


Figure 12. Le patron de conception MVC

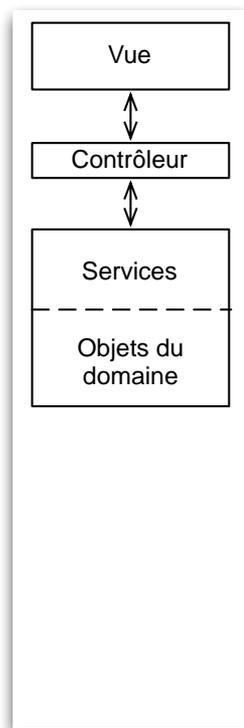


Figure 13. MVC - le découpage du modèle

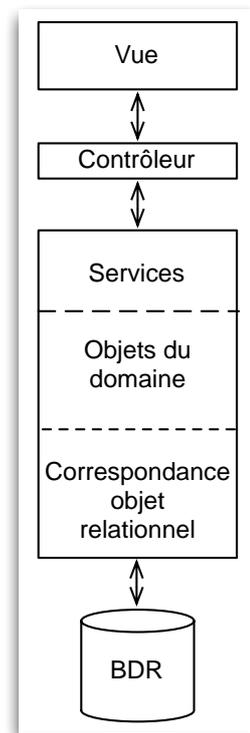


Figure 14. MVC - la correspondance objet relationnel

La vue est complètement dissociée du modèle. Cette dissociation permet d'envisager le support de différentes interfaces (Web, natives Android ou iOS, etc.) exploitant le même modèle. Dans une certaine mesure, des modifications dans l'implémentation du modèle peuvent ne pas impacter la vue. Le niveau de responsabilité "Contrôleur" est chargé d'assurer la communication entre la vue et le modèle en fonction des interactions de l'utilisateur.

Afin de maximiser le découplage de la vue et du modèle mais aussi le découplage entre les fonctions métier exposées à la vue et les implémentations de ces fonctions métier, le patron MVC traditionnel a évolué dans une approche dissociant deux aspects du modèle. La Figure 13 synthétise cette évolution de MVC : le modèle expose les fonctions métiers sous forme de services qui s'appuient sur les objets du domaine qui représentent les entités métiers.

Enfin, les applications d'entreprises stockent encore très majoritairement leurs données dans des bases de données relationnelles (BDR). Les BDR sont manipulables suivant une approche qui n'a rien à voir avec une orientation objet. Le patron MVC, dans le contexte du développement d'applications d'entreprise, a donc évolué pour prendre en compte cette distorsion entre le modèle de programmation des BDR et la programmation orientée objet. La Figure 14 synthétise la dernière version du patron que nous avons utilisé dans notre application. La couche de correspondance objet relationnel (COR) est chargée de fournir l'abstraction nécessaire au développeur de la couche modèle lui permettant de travailler selon une approche orientée objet. Notons que la spécification Java EE 1.5 (version précédent la version 6) a introduit la spécification Java Persistence API (JPA) afin de fournir une manière standard d'implémenter cette couche de correspondance objet relationnel.

Dans la suite du document nous utiliserons l'acronyme MVC-E (MVC Entreprise) pour désigner cette dernière variante du patron MVC.

5.2.3.2 L'injection de dépendances

L'injection de dépendances consiste à confier à un objet tiers (un conteneur) d'une part la gestion du cycle de vie des objets (création, "libération", etc.) et d'autre part la mise en relation des objets entre eux (l'injection des dépendances). La Figure 15 synthétise le principe de l'injection de dépendances dans un environnement Java.

La description des informations relatives au cycle de vie des objets et aux dépendances est effectuée via un ensemble de fichiers de configuration ou via des

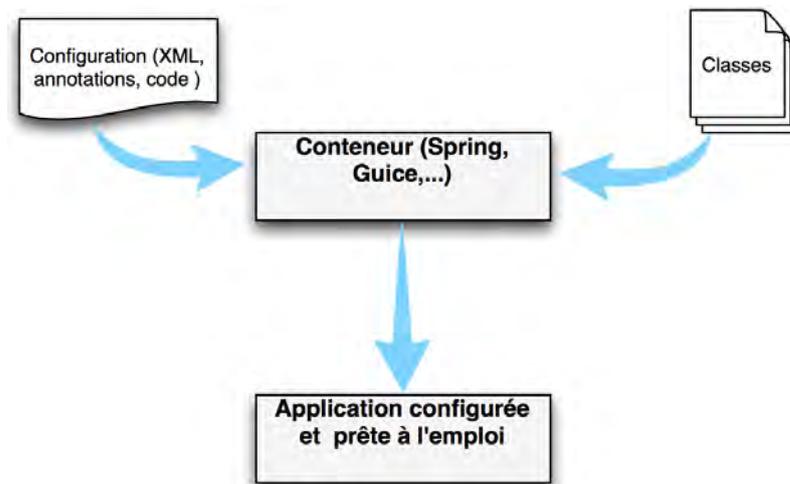


Figure 15. L'Injection de dépendances

annotations. Le conteneur, à partir de ces descriptions et des classes définies pour l'application, instancie les objets et met à jour les dépendances entre les objets à l'exécution, en général au lancement de l'application. Les conteneurs les plus populaires dans le monde Java sont Spring⁴ et Guice⁵.

Cette approche généralise l'approche conteneur de Java EE en l'étendant à tout type d'objets. Les conteneurs Java EE ne gère que certains types d'objets; par exemple, un conteneur Web gère essentiellement des objets de type Servlet.

L'injection de dépendances, en séparant les responsabilités entre objets en charge des traitements métiers et/ou applicatifs et objets responsables de la gestion du cycle de vie et des dépendances, favorise la modularité et la testabilité.

5.2.4 Le *framework* de développement

En développement informatique, le terme *framework* est utilisé pour désigner un ensemble de bibliothèques formant un tout cohérent pour le développement d'un type d'applications particulier. Java EE est victime de son succès et dispose donc d'un grand nombre de *frameworks* facilitant le développement d'applications Web.

Nous avons sélectionné notre framework Java EE de développement en appliquant la liste de critères suivante :

1. le framework devait faciliter la mise en œuvre des patrons de conception exposés dans la section 5.2.3;

4. <http://projects.spring.io/spring-framework/>

5. <https://github.com/google/guice>

2. le framework devait permettre un développement rapide de prototypes ;
3. le framework devait faciliter l'écriture des tests ;
4. le framework devait disposer d'une certaine visibilité sur le marché des frameworks de développement Web.

En 2005, un framework de développement d'applications Web en Ruby a fait son apparition dans le paysage du développement Web. Ruby on Rails⁶ (RoR) apporta dans le monde de la programmation Web un environnement extrêmement productif implémentant un grand nombre des meilleurs patrons de conception et proposant un environnement de test intégré à l'environnement de développement. De grandes compagnies comme Twitter ou Github ont adopté RoR. Le paradigme de développement RoR va dépasser les frontières de la sphère Ruby et dès 2006 apparaissent des tentatives de portage de l'approche RoR dans le monde Java EE. Grails⁷ est le framework Java EE s'étant imposé comme le framework incarnant le modèle RoR. Grails est Open Source et distribué sous licence Apache.

Grails fournit un environnement pour développer des applications Web implémentant le patron de conception MVC-E. La correspondance objet relationnel est mise en œuvre avec Hibernate⁸, une des implémentations les plus populaires de la spécification JPA (critère 1). Les différents services, contrôleurs et vues sont gérés à l'aide du conteneur Spring (critère 1). Fidèle à l'approche RoR, un dispositif d'"échafaudage" (*scaffolding* en anglais) permet la création rapide de vues et de contrôleurs à partir des objets du domaine (critère 2). Grails fournit un environnement de tests intégrés supportant nativement les bibliothèques de tests Java éprouvées (JUnit⁹ et Spock¹⁰) (critère 3). Enfin, Grails était jusqu'en 2014 un projet officiel de SpringSource, la filiale de VMware en charge de tous les projets autour de Spring. Grails est sorti en 2015 des projets SpringSource. Quand nous avons commencé les développements de la plateforme Tsaap en 2013, Grails répondait bien au critère 4.

6. <http://rubyonrails.org/>

7. <https://grails.org/>

8. <http://hibernate.org/>

9. <http://junit.org/>

10. <https://code.google.com/p/spock/>

5.2.5 Focus sur les interfaces clientes

Une des exigences formulées dans la section 5.2.1 est que l'application soit accessible aussi bien depuis un ordinateur portable que depuis un smartphone. Pour répondre à cette exigence, nous avons utilisé la boîte à outil CSS et Javascript Twitter Bootstrap¹¹ offrant la possibilité d'écrire un code d'interface commun s'adaptant, lors de l'exécution, au dispositif utilisé pour rendre l'interface. Les interfaces créées avec Twitter Bootstrap sont dites *responsive*. Comme son nom le suggère, Twitter Bootstrap est un outil conçu par Twitter ; il est Open Source.

5.2.6 Le système de gestion de bases de données relationnelles

Le système mis en œuvre a vocation à stocker une grande quantité d'informations : des informations sur les utilisateurs, des questions, des réponses, etc. Le système nécessite donc un dispositif de stockage de données. Nous avons opté pour un système de gestion de base de données relationnelles (SGBDR).

Nous n'avons pas de contrainte particulière sur le SGBDR en dehors du caractère Open Source. Nous avons fait le choix de MySQL¹². Notons que l'utilisation d'un outil de correspondance objet relationnel offre une couche d'abstraction permettant de changer à moindre coût de SGBDR, si cela s'avérait nécessaire.

5.3 Itération 1 - SVI, micro-blogging et prise de notes collaborative : Tsaap-Notes

5.3.1 Introduction

Le Tableau 6 présenté en fin de section 3.5 montre que la combinaison adéquate des qualités des SVI et des SEPT pourrait être un moyen de concevoir un système d'évaluation formative dépassant un grand nombre de limites des SVI. Notre démarche a donc consisté dans un premier temps à combiner dans une même plateforme le meilleur des SVI et le meilleur des SEPT utilisés dans les amphithéâtres. Cette démarche a conduit à une plateforme de micro-blogging

11. <http://getbootstrap.com/>

12. <https://www.mysql.fr/>

adaptée permettant :

1. une véritable prise de notes collaborative sur les supports de cours diffusés par l'enseignant pendant le cours ;
2. la pose d'une question interactive par l'enseignant sur une partie de son cours directement dans le dispositif utilisé par les étudiants ;
3. la prise de notes et les échanges suite à la présentation des résultats obtenus pour une réponse donnée.

Le système conçu, baptisé Tsaap-Notes, vient donc en support de stratégies ayant été identifiées comme relevant de l'évaluation formative : les apprenants deviennent des ressources d'instruction pour les autres apprenants. Cette itération est par ailleurs une première intégration d'interface de saisie de texte libre dans un dispositif offrant les fonctionnalités de SVI. Elle prépare le support de questions autorisant la saisie d'une réponse contenant un élément fourni en texte libre ; cette fonctionnalité fait l'objet de la troisième itération.

5.3.2 Impacts sur les processus

Tsaap-Notes vient donc en support du processus de Cours interactif. Sur les Figures 16 et 17, nous avons entouré en rouge les tâches ou activités sur lesquelles Tsaap-Notes intervient directement. Sur la Figure 17, la tâche de participation à la discussion a été entourée en pointillés. En effet, la discussion est a priori orale dans le processus, mais la présence d'un dispositif de micro-blogging utilisable pendant le cours n'exclut pas qu'une partie de la discussion se déroule via Tsaap-Notes.

5.3.3 Aperçu de Tsaap-Notes

Tsaap-Notes est une application Web. Pour utiliser l'application, il suffit de créer un compte sur la plateforme¹³. Chaque titulaire de compte a accès aux mêmes fonctionnalités : un utilisateur peut poster des notes, créer des *scopes* qui permettent d'agrèger un flux de notes, s'abonner à un ou plusieurs *scopes*. Quand un utilisateur s'abonne à un *scope*, il reçoit une notification lorsque de nouvelles notes ont été ajoutées au *scope*. La taille d'une note est limitée à 280 caractères. Une note peut contenir des *hashtags*¹⁴ qui sont utilisés pour qualifier et indexer la note. Un

13. <https://notes.tsaap.eu/tsaap-notes>

14. Un *hashtag* est un mot préfixé par le caractère "#".

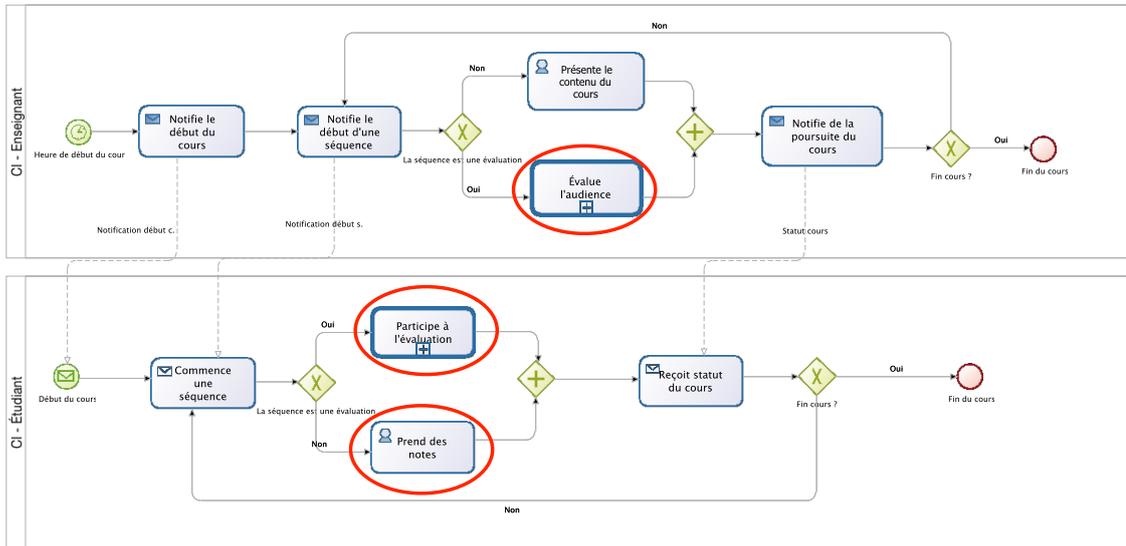


Figure 16. Impact de Tsaap-Notes sur le processus de Cours Interactif

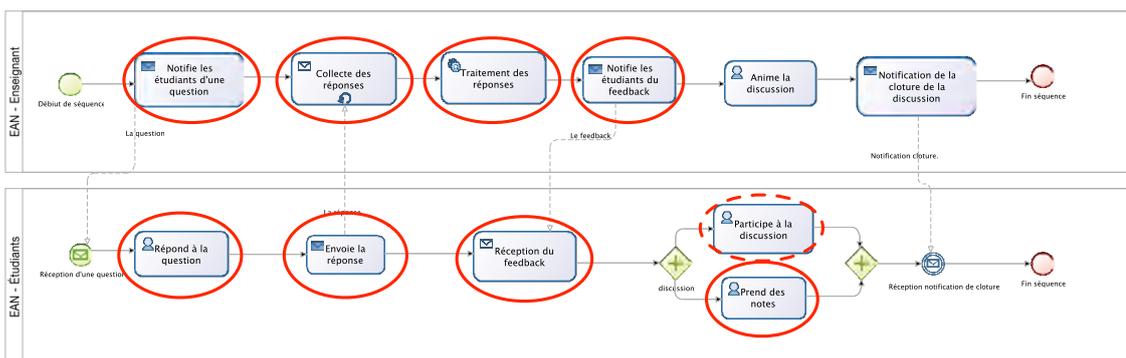


Figure 17. Impact de Tsaap-Notes sur le processus d'évaluation d'audience nominale



Figure 18. L'interface de Tsaap-Notes dans un smartphone

utilisateur peut marquer une note comme "favorite" et peut répondre à une note particulière. La distinction entre un enseignant et un étudiant est faite quand un utilisateur crée un *scope* : l'utilisateur doit indiquer s'il est enseignant pour le *scope* qu'il vient de créer. Un utilisateur peut utiliser Tsaap-Notes avec son smartphone, son ordinateur ou tout autre appareil qui est connecté à Internet. La Figure 18 montre l'interface de Tsaap-Notes à partir d'un smartphone.

5.3.4 Lier les ressources du cours avec les notes

Les plateformes de micro-blogging traditionnelles ne sont pas conçues pour annoter des ressources. Tout au plus il est possible de renseigner dans un message une URL ou un hashtag référençant respectivement une ressource ou un événement. Le concept de *scope* a été introduit dans Tsaap-Notes pour prendre en charge de manière transparente le double lien « physique » et sémantique entre une ressource présentée pendant le cours et un flux de notes relatif à la ressource.

Un *scope* est caractérisé par un nom, une description pouvant contenir des *hashtags* et l'URL de la ressource cible, généralement une ressource présentée pendant le cours. Une note dans Tsaap-Notes peut être liée à un *scope*. Dans ce cas, la note hérite des hashtags spécifiés dans la description du *scope*. La Figure 19



Figure 19. Exemples de *scopes*

montre deux exemples de *scope* dans l'interface de Tsaap-Notes.

Un "fragment" est un *hashtag* technique (non sémantique) utilisé pour identifier un fragment de la ressource référencée par un *scope*. Une note liée à un *scope* peut également être liée à un "fragment", i.e. à une sous partie de la ressource utilisée pendant le cours. Afin de permettre l'annotation des ressources utilisées pendant le cours qui est préconisée dans Kiewra *et al.* (1988), nous avons mis en place dans Tsaap-Notes la fonction permettant l'intégration dans une page Web de la liste des notes liées à un *scope* et à un « fragment ». Ensuite, en parallèle de Tsaap-Notes, nous avons développé une solution reposant sur le standard OASIS DocBook¹⁵ permettant d'obtenir les diapositives du support de cours au format HTML. Une feuille de style transforme le XML Dockbook en HTML et génère un site Web contenant le diaporama et la version intégrée de Tsaap-Notes affichant les notes liées à la diapositive en cours. Dans ce cas d'utilisation, l'URL du *scope* correspond à l'URL du diaporama et chaque diapositive est associée à un fragment. La Figure 20 montre la vue intégrée de Tsaap-Notes dans une page Web contenant une diapositive utilisée lors d'un cours. La première zone délimitée en rouge est une barre de navigation permettant aux étudiants de naviguer sur les diapositives présentées durant le cours. La deuxième zone affiche le contenu de la diapositive en cours. La troisième zone est l'interface intégrée de Tsaap-Notes permettant à l'étudiant d'ajouter et de consulter les notes liées à la diapositive en cours.

Le triptyque "fragment - *scope* - vue intégrée" est la solution mise en œuvre dans Tsaap-Notes pour relier de manière transparente « physiquement » et sémantiquement une ressource utilisée pendant le cours avec les notes qui lui sont relatives. Cette liaison automatique entre les notes et les diapositives permet aux étudiants d'annoter les supports de cours.

15. <http://www.docbook.org>

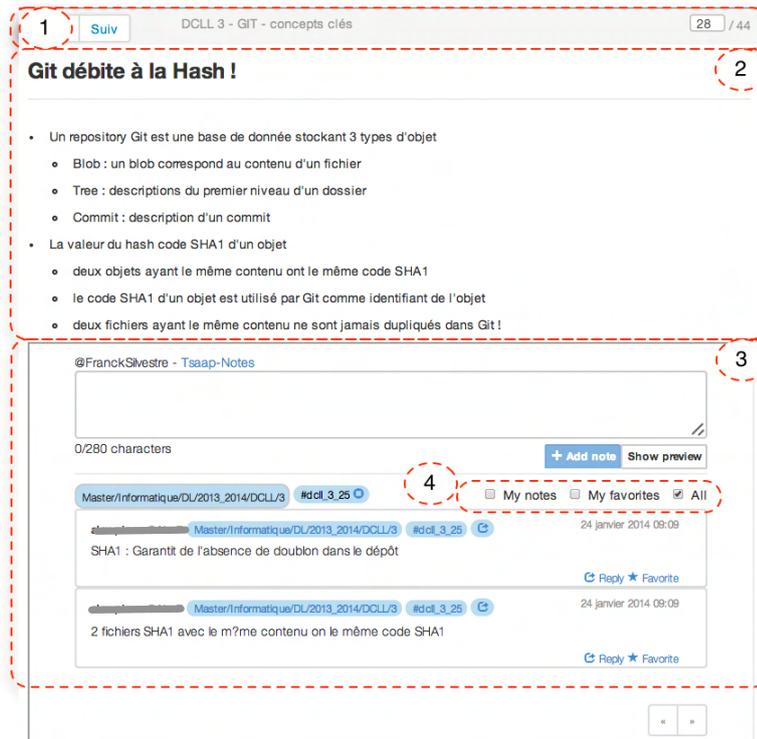


Figure 20. Annotation d'une diapositive avec Tsaap-Notes

5.3.5 Les vues personnalisées réutilisables

Nous avons exposé précédemment que les fonctions de personnalisation offertes par les plateformes de micro-blogging actuelles sont intéressantes dans la perspective de l'apprentissage. Cependant, ces fonctions ne permettent pas aux étudiants d'accéder facilement à leurs propres filtres sur les différents contenus fournis par tous les contributeurs dans le contexte d'un cours. Tsaap-Notes propose à chaque utilisateur accédant à une liste de notes la possibilité de filtrer la liste afin qu'elle ne contienne que les notes les plus pertinentes pour lui. Sur la Figure 20, la quatrième zone rouge contient l'interface permettant d'activer la fonction de filtre où l'utilisateur peut activer ou désactiver trois options d'affichage pour une diapositive donnée :

- l'affichage de toutes les notes liées à la diapositive,
- l'affichage des notes écrites par l'utilisateur,
- l'affichage des notes que l'utilisateur a marqué comme favorites.

Les différentes combinaisons de ces options permettent par exemple à l'étudiant dans un premier temps de consulter toutes les contributions et de sélectionner celles



Figure 21. Interface d'édition d'une question

qui sont les plus pertinentes pour lui (en les marquant comme favorites), puis pendant une préparation d'examen de se concentrer sur la liste des notes dont il est l'auteur et sur celles qu'il a sélectionnées.

5.3.6 Le système de votes interactifs

Pour doter Tsaap-Notes de la fonction de système de votes interactifs, nous avons mis en œuvre la fonction que nous désignons par l'expression "*note as a question*" : l'enseignant peut saisir le contenu d'une note au format Moodle Gift¹⁶ pour décrire une question interactive. Ensuite, le moteur de Tsaap-Notes est en mesure de proposer, à la demande de l'enseignant, l'interface permettant aux étudiants de répondre à la question. Le moteur Tsaap-Note, toujours à la demande de l'enseignant, présente une vue graphique instantanée des résultats. La vue des résultats est partagée avec l'ensemble des étudiants. La Figure 21 donne un exemple d'édition de question. L'ajout d'une question s'appuie sur l'interface d'ajout de note. Lorsque la note est ajoutée, elle est traitée par Tsaap-Notes pour générer la vue correspondant à une question si le texte entré par l'enseignant est un texte conforme avec le format Moodle Gift. L'auteur de la question peut contrôler le début et l'arrêt de la période de soumission de réponses.

La Figure 22 montre l'interface de soumission proposée à tous les autres utilisateurs. Les étudiants soumettent leurs réponses au système en utilisant les éléments d'interaction offerts par l'interface (boutons radio, cases à cocher, bouton Soumettre).

16. https://docs.moodle.org/2x/fr/Format_GIFT

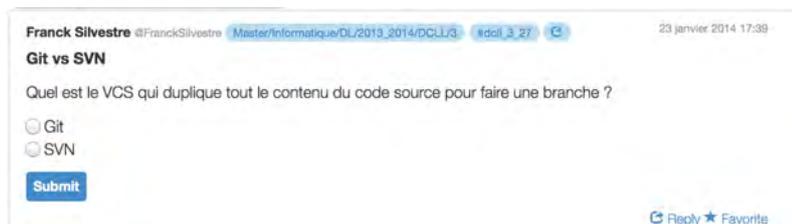


Figure 22. Interface de soumission d'une réponse



Figure 23. Présentation des résultats

La Figure 23 présente l'affichage des résultats lorsque l'enseignant ferme la période de soumission.

5.3.7 Synthèse des résultats

Tsaap-Notes, dès sa première version, tire parti des qualités clés des systèmes de votes interactifs :

- support d'activités de référence correspondant aux critères A-1 et A-2 ;
- support de stratégies correspondant aux critères S1, S-2.1, S-2.2, S3, S4 et S5.

En combinant les fonctionnalités des systèmes de prises de notes collaboratives aux fonctionnalités de SVI, Tsaap-Notes introduit de nouvelles interactions autorisant la saisie de texte libre (critère I-ST) et de nouveaux dispositifs permettant aux étudiants d'être des ressources d'instruction pour les autres (critère S-4). D'autre part, les fonctionnalités de micro-blogging proposées par Tsaap-Notes permettent aux étudiants de fournir des réponses aux contributions des autres ; ces réponses représentent ainsi une première forme de feedback textuel (critère I-TF). Enfin, les étudiants bénéficient de la possibilité de personnaliser les vues sur les différentes notes prises pendant le cours favorisant ainsi la responsabilité de leur propre

apprentissage (critère S-5).

5.4 Itération 2 - La génération semi-automatique de tests de révisions

5.4.1 Introduction

L'utilisation du format Moodle Gift facilite le passage de Tsaap-Notes à Moodle et vice versa. Un enseignant stockant ses questions au format Gift dans un fichier texte peut injecter les questions posées pendant le cours dans une plate-forme Moodle. Les étudiants peuvent alors rejouer les questions après le cours, par exemple pour préparer leurs examens.

Cette approche est un premier pas pour faciliter l'utilisation de tests sommatifs à visée formative. En effet, Moodle est une plateforme utilisée pour faire passer des tests sommatifs aux étudiants. Faciliter la réutilisation des questions posées pendant le cours dans Moodle pour la constitution de tests sommatifs contribue à développer une activité relevant de l'évaluation formative.

La deuxième itération dans la conception et la mise en œuvre de la plateforme Tsaap a pour ambition de franchir une avancée décisive pour faciliter la réutilisation des questions posées en cours.

Différents travaux (Ricketts et Wilks, 2002; Miller, 2009) montrent que les évaluations informatisées facilitent la mise en œuvre d'évaluations formatives fréquentes pour des cohortes comprenant un grand nombre d'étudiants. La qualité du feedback proposé aux étudiants et relatif à leurs travaux ou à leurs résultats d'évaluation est une clé de succès pour l'amélioration de leur apprentissage (Black et Wiliam, 1998; Hattie et Jaeger, 1998; Higgins *et al.*, 2002). Nicol et Macfarlane-Dick (2006) 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.

Les expériences décrites dans 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 feedback 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.

Les situations d'auto-évaluations informatisées posent aussi la question du feedback proposé à l'étudiant. 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 (IMS, 2012)), 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 (telles 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 et 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 » (Higgins, 2000; Chanock, 2000), 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 et McKenna, 2003).

Nous proposons d'aborder ces différents problèmes soulevés par les auto-évaluations informatisées à l'aide de l'approche baptisée *Notes as Feedback* (NaF) qui consiste à réutiliser les notes prises par les étudiants pendant le cours comme source de feedback lorsqu'ils rejouent les questions en vue d'effectuer une auto-évaluation. Cette approche (1) dispense l'enseignant de saisir le feedback (tout en maintenant cet acteur comme source indirecte du feedback, à travers les idées transmises durant le cours et retranscrites par les étudiants sous forme de notes), (2) délivre aux étudiants un feedback spécifique qui correspond aux productions réalisées durant le cours auquel ils participent (si un enseignant dispense le même enseignement à différentes promotions d'étudiants, le feedback sera différent d'une promotion à une autre), (3) garantit la compréhension du feedback par les étudiants puisqu'il résulte de leurs propres notes rédigées dans leur « dialecte », et (4) permet la corrélation directe de la question et du feedback associé avec la structure du cours et son déroulement, puisque les questions sont posées par l'enseignant tout au long

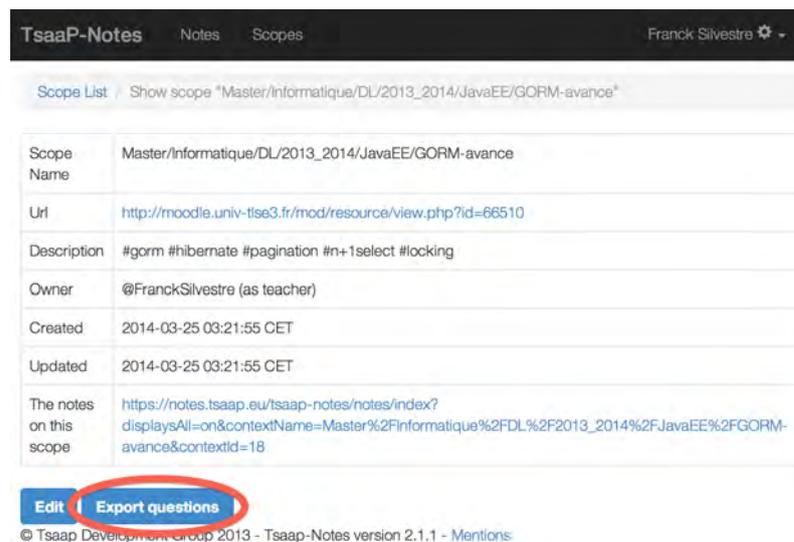


Figure 24. La fonction d'exportation des questions relatives à un *scope*

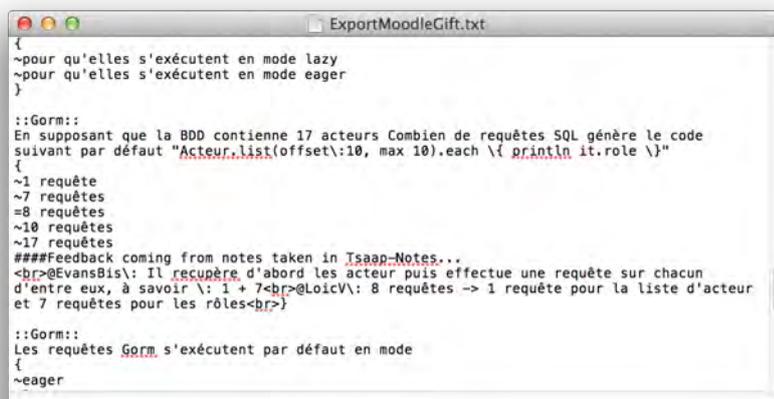
de l'apprentissage des concepts enseignés.

5.4.2 Tsaap-Notes et la génération de tests d'auto-évaluation

Afin de faciliter la constitution d'un test d'auto-évaluation à partir des questions interactives proposées durant le déroulement du cours, nous avons conçu et développé une fonction d'exportation des questions au sein de Tsaap-Notes.

Le propriétaire d'un *scope* (c'est-à-dire un enseignant) dispose d'un bouton permettant d'exporter toutes les questions interactives liées au *scope*; la Figure 24 présente l'interface de Tsaap-Notes incluant ce bouton d'exportation. Le moteur d'export recherche, pour chaque question du *scope*, les notes prises par les étudiants sur la question en cours de traitement et injecte ces notes comme feedback général de la question.

L'exportation est aujourd'hui conforme au format Moodle Gift. Le fichier généré par Tsaap-Notes (Figure 25) est donc directement importable dans Moodle à l'aide de la fonctionnalité d'importation de questions permettant d'alimenter la banque de questions spécifique à un cours Moodle. Les questions importées peuvent être éditées dans Moodle; la Figure 26 présente le formulaire d'édition d'une question issue de Tsaap-Notes dans Moodle. Le champ « Feedback général » est ici automatiquement renseigné avec les notes prises par les étudiants durant le



```
{
~pour qu'elles s'exécutent en mode lazy
~pour qu'elles s'exécutent en mode eager
}

::Gorm:
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 \}"
{
~1 requête
~7 requêtes
=8 requêtes
~10 requêtes
~17 requêtes
###Feedback coming from notes taken in Tsaap-Notes...
<br>@EvansBis\: Il récupère d'abord les acteurs puis effectue une requête sur chacun
d'entre eux, à savoir \: 1 + 7<br>@LoicV\: 8 requêtes -> 1 requête pour la liste d'acteur
et 7 requêtes pour les rôles<br>}

::Gorm:
Les requêtes Gorm s'exécutent par défaut en mode
{
~eager
}
```

Figure 25. Le fichier généré par l'export des questions

cours dispensé en face-à-face. L'enseignant dispose donc s'il le souhaite d'un moyen d'adapter le feedback associé à une question. Le test d'auto-évaluation peut alors être créé par l'enseignant en sélectionnant, dans la banque de questions, celles issues de l'importation du fichier Tsaap-Notes. Les questions accompagnées de leur feedback général sont donc créées de manière automatique à l'importation du fichier généré par Tsaap-Notes ; la création du test agrégeant ces questions reste à ce jour manuelle.

Lorsque les étudiants réalisent un test d'auto-évaluation, ils disposent, pour chacune des questions, de leur score, de la réponse correcte, et des notes associées à la question qui ont été formulées par les étudiants durant le déroulement du cours en face-à-face ; la Figure 27 illustre l'interface Moodle affichant le résultat d'un étudiant sur une question d'un test d'auto-évaluation.

5.4.3 Synthèse des résultats

En s'appuyant sur la réutilisation de questions et de notes créées dans Tsaap-Notes, le dispositif « Notes as Feedback » dispense l'enseignant de la saisie du feedback textuel pour chaque question (celui-ci est spécifié de manière automatique par l'injection des notes prises par les étudiants). Le dispositif facilite donc l'introduction d'une activité de révision supplémentaire fondée sur des évaluations informatisées. Ces évaluations permettent aux étudiants de travailler aux périodes, fréquences et lieux de leur choix. Les feedbacks proposés dans ces tests sont des feedbacks textuels provenant de la saisie de texte libre des étudiants. Si les questions

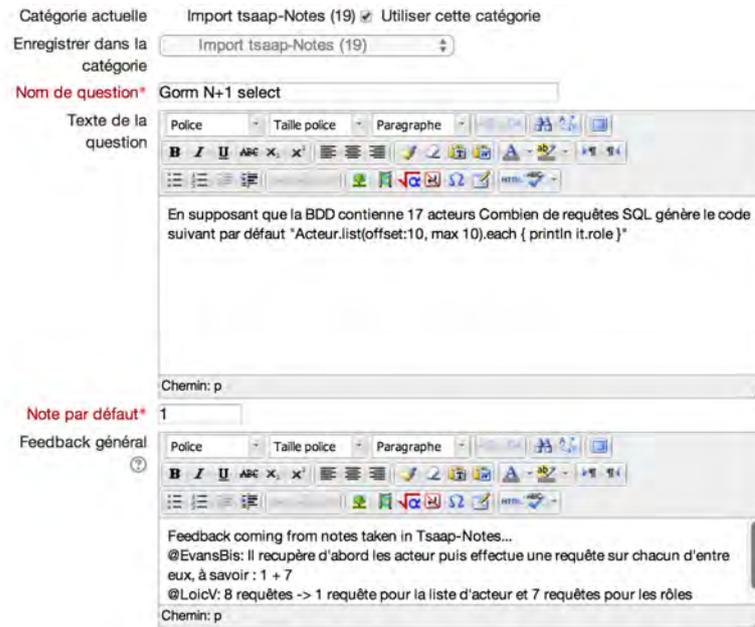


Figure 26. Édition d'une question issue de Tsaap-Notes et importée dans Moodle

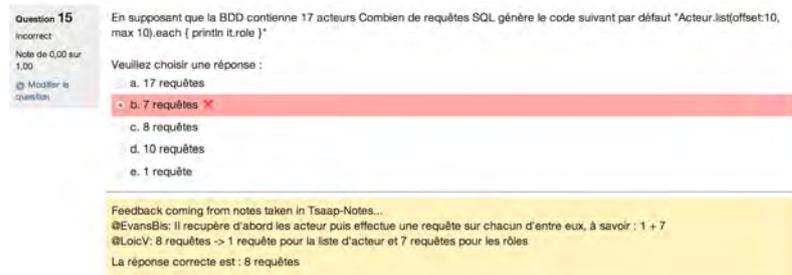


Figure 27. Affichage du feedback dans Moodle pour une réponse donnée par un étudiant

sont elles-mêmes ré-utilisées dans le contexte d'évaluations sommatives, Tsaap-Note apparait de fait comme un outil facilitant l'utilisation de tests sommatifs à visée formative (critère A-5) : les questions des tests sommatifs sont les questions utilisées en premier lieu dans les évaluations d'audience menées avec Tsaap-Notes puis dans les tests de révisions générés de manière semi-automatique par Tsaap-Notes. L'approche « Notes as Feedback » valorise l'engagement des étudiants durant le cours en face à face en recyclant leur prise de notes au sein d'une activité initiée par l'enseignant ; elle renforce la stratégie consistant à utiliser les étudiants comme ressources d'instruction pour les autres (critère S-4).

5.5 Itération 3 - Le SEA Tsaap

5.5.1 Introduction

Au terme de la deuxième itération, Tsaap-Notes est une plateforme combinant habilement les fonctionnalités de système de votes interactifs et de système d'échange et de partage de textes, tout en facilitant l'utilisation de tests sommatifs à visée formative. Cependant Tsaap-Notes est loin de lever l'ensemble des verrous identifiés dans la section 3.6 :

1. les interactions permettant de répondre aux questions posées via Tsaap-Notes restent celles d'un système de votes interactifs standard ; Tsaap-Notes ne permet donc toujours pas la prise en compte de questions dont les réponses contiennent des éléments textuels saisis librement ;
2. le feedback renvoyé par le SVI Tsaap-Notes reste un indicateur numérique ;
3. la confrontation d'idées entre pairs n'est pas suscitée explicitement par Tsaap-Notes ; cette confrontation reste dépendante du bon vouloir de l'enseignant ou des étudiants (qui peuvent éventuellement utiliser Tsaap-Notes pour communiquer pendant le cours) ;
4. les activités d'auto-évaluation et d'évaluation par les pairs ne sont pas supportées explicitement par Tsaap-Notes.

L'objectif de la troisième itération est de traiter ces différents verrous en implantant le processus de soumission de réponse en N phases décrit dans la section 4.7.



@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 Learn

Figure 28. Soumission d'une réponse : première phase

5.5.2 La mise en œuvre du processus en N phases

Cette section présente les phases du processus en N-phases sous l'angle de son implantation dans Tsaap-Notes.

La première phase (première soumission de réponse) consiste à proposer à l'étudiant un formulaire permettant (1) de répondre à la question à choix multiples, (2) de saisir l'explication justifiant son choix (saisie de texte libre), et (3) d'indiquer son degré de confiance en sa réponse. La Figure 28 présente le formulaire tel qu'il a été implémenté dans Tsaap-Notes. Dès cette première étape, l'étudiant est mis en situation d'apprentissage réflexif : en indiquant le degré de confiance en sa réponse, l'étudiant est amené à réfléchir sur ses propres connaissances et/ou compétences. Cet indice est un des éléments clés de l'algorithme sous-tendant la deuxième phase.

La deuxième phase de soumission (le conflit socio-cognitif et la re-soumission de réponse) consiste à présenter à l'étudiant le même formulaire, mais cette fois pré-rempli avec les éléments de sa première réponse, et accompagné d'une réponse différente de la sienne qui a été fournie par un autre étudiant participant à la même session interactive. La Figure 29 illustre l'interface correspondant à cette phase au sein de Tsaap-Notes. Notons qu'aucune indication sur l'identité de l'auteur de la réponse alternative n'est présentée à l'étudiant afin d'écartier tout phénomène parasite d'influence relationnelle.

La troisième phase (évaluations par les pairs et feedback) illustrée par la Figure 30 présente les résultats obtenus par l'ensemble des étudiants ayant participé au questionnaire (la barre verte indique le taux de bonnes réponses, la barre rouge

The screenshot shows a forum post by @FranckSilvestre. The question is: "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 ?". The user has selected "Non" and provided an explanation: "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." The confidence scale is set to 3. A red box highlights the alternative response: "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." Another red box highlights the submission form with the text: "Second formulaire de soumission : une chance de modifier la première réponse".

Figure 29. Soumission d'une réponse : deuxième phase

indique le taux de mauvaises réponses); chaque étudiant prend connaissance du score qu'il a obtenu à la question en cours ; il peut se situer par rapport à l'ensemble des étudiants en consultant les taux de bonnes et mauvaises réponses. Dans cette même interface, chaque étudiant est invité à évaluer, à l'aide d'une échelle de Lickert graduée de 1 à 5, trois explications distinctes correspondant chacune à la bonne réponse. Durant cette phase, chaque étudiant est en situation d'évaluateur des productions provenant de ses pairs, et chaque explication correspondant à une réponse correcte est évaluée au moins trois fois. Une fois cette phase terminée, Tsaap-Notes calcule la note moyenne de chaque explication évaluée et présente les résultats du processus d'évaluation sous la forme d'une liste ordonnée par ordre décroissant des explications évaluées.

La quatrième phase (discussion) est animée par l'enseignant. Les étudiants peuvent prendre des notes à l'aide de Tsaap-Notes au cours de la discussion.

5.5.3 Focus sur l'algorithme d'association de réponses pour le conflit socio-cognitif

La phase 2 du processus en N phases, dont l'objectif est de générer un conflit socio-cognitif, nécessite la mise en œuvre d'un algorithme capable d'associer de

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

Results **La licence GPL et Javascript** - (response count : 4)

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 75%

Non 25%

Do not know

Your score : 0.0%

A last work waits for you: please give a grade to the explanations given for the good answer (1: "not usefull" to 5: "very usefull").

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.
4

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.
5

Sous licence GPL, Cecill (equivalent pour le droit français), toute modification/extension doit être distribuée sous licence compatible;
2

Submit

Figure 30. Soumission d'une réponse : troisième phase

manière pertinente deux réponses obtenues à l'issue de la phase 1. Le principe sous-tendant notre algorithme consiste alors à associer une réponse a_i à une réponse a_j lorsque les deux réponses ont fait l'objet de choix distincts et que l'explication associée à a_i est de taille raisonnable ; si plusieurs réponses peuvent être associées à a_j , l'algorithme privilégie la diversité dans l'association des réponses afin d'exposer des associations différentes aux apprenants, tout en favorisant les réponses ayant le plus fort degré de confiance. Afin de présenter notre algorithme, nous proposons dans ce qui suit une formalisation du problème et des éléments de la solution retenue.

Dans la suite du document, q désigne une question à choix multiples.

Définition 3. Une réponse a à une question à choix multiples q est un triplet

$$a = (s; d; e)$$

où

- s est le score calculé en fonction de la réponse de l'étudiant ;
- d est le degré de confiance fourni par l'étudiant sur sa réponse ;
- e est le nombre de caractères composant l'explication fournie par l'étudiant.

Propriété 1. Soit A l'ensemble des réponses obtenues à q . La relation \geq sur A définie par

$$\forall (a_i = (s_i; d_i; e_i), a_j = (s_j; d_j; e_j)) \in A^2, a_i \geq a_j \Leftrightarrow d_i \geq d_j$$

est une relation d'ordre sur A .

Définition 4. L'ensemble A_c des réponses correctes à q est défini comme suit

$$A_c = \{a_i = (s_i; d_i; e_i) \in A \mid s_i = \text{MaxScore}\}$$

où MaxScore est le score maximum qu'un étudiant peut obtenir à une réponse q .

Définition 5. L'ensemble $\overline{A_c}$ des réponses incorrectes à q est le complémentaire A_c dans A .

Définition 6. L'ensemble des réponses à q pouvant être proposées pour constituer un conflit socio-cognitif est l'ensemble

$$C = \{a_i = (s_i; d_i; e_i) \in A \mid e_i > \text{TailleMinExp}\}$$

où $TailleMinExp$ est la taille minimum requise d'une explication pour qu'une réponse soit considérée comme associable lorsqu'il s'agit de générer un conflit socio-cognitif.

Définition 7. L'ensemble des réponses correctes à q pouvant être proposées pour générer un conflit sociocognitif est l'ensemble

$$C_c = C \cap A_c$$

Définition 8. L'ensemble \overline{C}_c des réponses incorrectes à q pouvant être proposées pour constituer un conflit socio-cognitif est le complémentaire de C_c dans C .

Définition 9. Soit $A_c = \{a_1, a_2, a_3, \dots, a_n\}$ où $\forall(i, j) \in \{1, \dots, n\} \quad i < j \Rightarrow a_i \geq a_j$ l'ensemble ordonné des réponses correctes à q .

Soit $\overline{C}_c = \{c_1, c_2, c_3, \dots, c_p\}$ où $\forall(i, j) \in \{1, \dots, p\} \quad i < j \Rightarrow c_i \geq c_j$ l'ensemble ordonné des réponses incorrectes à q pouvant être proposées pour constituer un conflit.

La fonction $f : A_c \rightarrow \overline{C}_c$ est la fonction associant à tout élément $a_i \in A_c$ l'élément

$$f(a_i) = c_{i \bmod p}$$

où $i \bmod p$ est le reste de la division euclidienne de i par p .

Définition 10. Soit $\overline{A}_c = \{a_1, a_2, a_3, \dots, a_n\}$ où $\forall(i, j) \in \{1, \dots, n\} \quad i < j \Rightarrow a_i \geq a_j$ l'ensemble ordonné des réponses incorrectes à q .

Soit $C_c = \{c_1, c_2, c_3, \dots, c_p\}$ où $\forall(i, j) \in \{1, \dots, p\} \quad i < j \Rightarrow c_i \geq c_j$ l'ensemble ordonné des réponses correctes à q pouvant être proposées pour constituer un conflit.

La fonction $g : \overline{A}_c \rightarrow C_c$ est la fonction associant à tout élément $a_i \in \overline{A}_c$ l'élément

$$g(a_i) = c_{i \bmod p}$$

où $i \bmod p$ est le reste de la division euclidienne de i par p .

À partir de ces définitions, la Figure 31 illustre les différentes étapes de l'algorithme développé dans Tsaap-Notes. La première étape prend en entrée l'ensemble A et crée les 4 ensembles A_c , \overline{A}_c , C_c et \overline{C}_c . Ces ensembles sont implémentés sous la forme de listes dans lesquelles les éléments sont ordonnés selon la relation d'ordre introduite par la Propriété 1. Puisque dans l'implémentation actuelle de

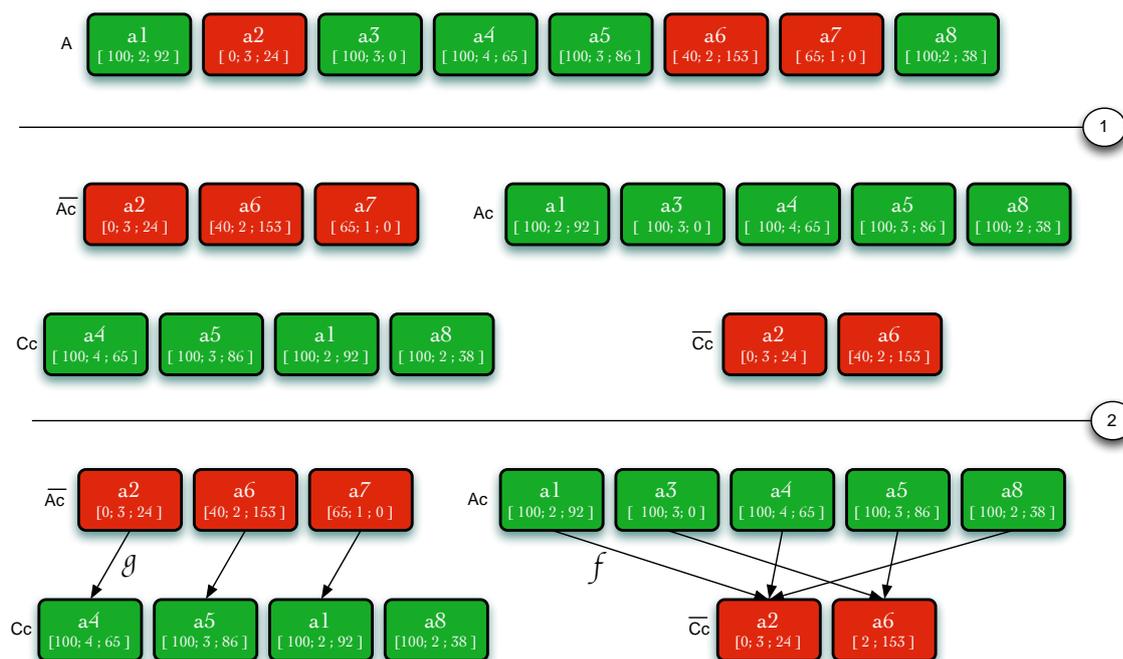


Figure 31. L’algorithme pour générer le conflit socio-cognitif

Tsaap-Notes le score maximum qui peut être obtenu à une question est toujours 100, le paramètre *MaxScore* a été fixé à cette même valeur. Le paramètre *TailleMinExp* correspondant à la taille minimum requise d’une explication pour qu’une réponse soit considérée comme associable pour générer un conflit socio-cognitif ; il a été fixé à 10. La deuxième étape de l’algorithme récupère dans un tableau associatif le résultat de l’application de la fonctions f (respectivement g) sur tous les éléments de l’ensemble A_c (respectivement $\overline{A_c}$).

5.5.4 La génération semi-automatique de tests améliorée

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 ; la Figure 32 montre l’interface proposée par Moodle à l’étudiant effectuant une relecture de sa copie.

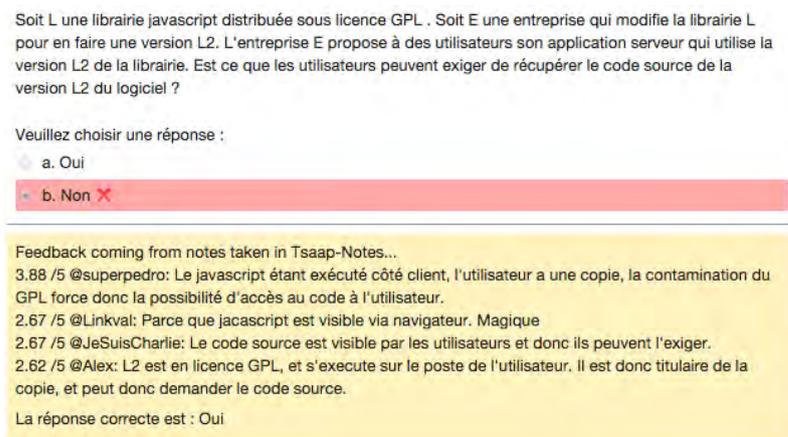


Figure 32. L'interface de Moodle présentant le feedback provenant de Tsaap

5.5.5 Synthèse des résultats

Le processus en N phases permet d'intégrer au sein de Tsaap-Notes d'une part de nouvelles formes d'interaction telles que la saisie de texte libre (critère I-ST) ou la fourniture de feedbacks textuels (critère I-FT), et d'autre part l'évaluation par les pairs (critère A-4). Les stratégies de conception de situations de discussions en classes (critère S-2.1) et d'utilisation des étudiants comme ressources d'instruction pour les autres (critère S-4) sont renforcées. L'activité d'évaluation à l'aide de commentaires textuels (critère A-3) est partiellement couverte par l'introduction de feedbacks textuels. À travers certaines tâches proposées comme l'évaluation par l'étudiant du niveau de confiance dans la réponse qu'il fournit, Tsaap-Notes renforce le fait que les apprenants deviennent responsables de leur propre apprentissage (critère S-5).

5.6 Conclusion

Trois jalons ont marqué la conception et la mise en œuvre de notre système d'évaluation d'audience. Au terme de chacun de ces jalons, Tsaap-Notes a intégré ou renforcé le support d'activités de référence ou de stratégies caractéristiques de l'évaluation formative au delà de ce que proposent les SVI actuels.

Le tableau 8 propose une première évaluation de la plateforme Tsaap-Notes à l'aide de la grille proposée dans la section 3.2.2. Tsaap-Notes remplit la quasi-totalité des critères du système cible spécifié dans la section 3.5. Le critère mesurant la capacité à engager les étudiants reste inconnu à ce stade. À l'issue de chaque

jalón, une expérimentation a été menée ayant pour objectif principal la mesure de l'engagement des étudiants dans les nouvelles activités introduites par notre approche. La troisième partie du manuscrit est consacrée à la présentation et à l'analyse des résultats d'expérimentations.

Table 8. 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															?

Troisième partie
Expérimentations

Chapitre 6

Contexte général des expérimentations

Sommaire

6.1	Contexte humain	113
6.2	Contexte technique	114
6.2.1	Le réseau de l'Université	114
6.2.2	L'environnement d'exploitation	114
6.3	Validité des expérimentations	114

6.1 Contexte humain

La plateforme Tsaap-Notes a été testée à l'Université Toulouse III Paul Sabatier dans le contexte de cours de première année de Master Informatique spécialité Développement Logiciel. Trois expérimentations relatives aux trois itérations ont été menées durant trois années universitaires consécutives : 2012-2013, 2013-2014 et 2014-2015. Les groupes comprenaient entre 30 et 50 étudiants.

Vingt tablettes Android ont été achetées par la formation pour équiper les étudiants ne disposant pas de matériel informatique permettant la connexion Internet en amphithéâtre. Les deux premières années, une quinzaine d'étudiants empruntait une tablette. Durant la troisième année, seule trois tablettes ont été empruntées et près d'un tiers des étudiants ont utilisé leur smartphone pour les activités menées avec Tsaap-Notes.

Table 9. Éléments clés de l’environnement d’exploitation

Environnement	Produit	Version
Système d’exploitation	Linux Cent OS ¹	6
Serveur d’application Java EE	Apache Tomcat ²	7
Serveur de bases de données	MySQL ³	5
Runtime Java	Oracle JDK ⁴	7

Chaque année, une vingtaine de minutes au maximum ont été consacrées à la présentation de Tsaap-Notes aux étudiants participant à l’expérimentation.

6.2 Contexte technique

6.2.1 Le réseau de l’Université

Les étudiants et enseignants utilisaient le réseau WIFI *eduroam*⁵ fourni par l’Université. Au cours des expérimentations, le réseau n’a pas fait défaut et la bande passante offerte fut confortable pour l’usage de Tsaap-Notes en amphithéâtre.

6.2.2 L’environnement d’exploitation

La plateforme Tsaap-Notes est déployée sur un serveur virtualisé disposant d’un cœur CPU 64 bits et de 2 Go de RAM.

Les autres éléments clés de l’environnement d’exploitation sont décrits dans le Tableau 9. Le serveur d’application et le serveur de bases de données sont déployés sur le même serveur. Une seule instance de Tomcat est déployée pour héberger la plateforme Tsaap-Notes.

6.3 Validité des expérimentations

Les expérimentations qui ont été menées ont deux biais notables :

- l’absence de groupes de contrôle permettant de comparer certains résultats au regard d’un groupe n’ayant pas utilisé la plateforme Tsaap-Notes ;

1. <https://www.centos.org/>

2. <http://tomcat.apache.org/>

3. <https://www.mysql.fr/>

4. <http://www.oracle.com/technetwork/java/javase/downloads/index.html>

5. <http://www.eduroam.fr/>

- le fait que le concepteur de la plateforme soit l'enseignant qui menait les expérimentations.

Ces biais nécessitent de considérer certains des résultats obtenus avec précaution et incitent à mener des expérimentations plus rigoureuses (avec groupes de contrôles, différentes matières, différents enseignants, différents niveaux, etc.) afin de confirmer les résultats encourageants constatés dans nos expériences.

Chapitre 7

Les résultats d'expérimentations

Sommaire

7.1	Expérimentation suite à l'itération 1	118
7.1.1	Objectifs	118
7.1.2	Contexte	118
7.1.3	Résultats quantitatifs	119
7.1.4	Résultats qualitatifs	119
7.1.5	Synthèse des résultats	120
7.2	Expérimentation suite à l'itération 2	121
7.2.1	Objectifs	121
7.2.2	Contexte	121
7.2.3	Participation à la prise de notes sur les questions interactives	121
7.2.4	Participation des étudiants aux tests de révision	122
7.2.5	Résultats au test d'évaluation sommative	123
7.2.6	Synthèse des résultats	123
7.3	Expérimentation suite à l'itération 3	124
7.3.1	Objectifs	124
7.3.2	Contexte	124
7.3.3	Résultats quantitatifs	125
7.3.4	Résultats qualitatifs	128
7.3.5	Synthèse des résultats	129
7.4	Conclusion	129

7.1 Expérimentation suite à l'itération 1

7.1.1 Objectifs

La version de Tsaap-Notes obtenue à l'issue de la première itération est une plateforme combinant les fonctionnalités d'un système de votes interactifs (SVI) avec les fonctionnalités d'un système d'échange et de partage de textes (SEPT). Cette combinaison permet d'introduire dans le SVI Tsaap-Notes des fonctionnalités de saisie de textes libres associées à la pose de questions interactives : les étudiants peuvent prendre des notes sur les questions posées suite à la présentation des résultats.

L'expérimentation menée avait pour objectif de valider les hypothèses suivantes :

- le SVI proposé par Tsaap-Notes engage les étudiants de manière conforme aux données fournies par la littérature scientifique ;
- le couplage des fonctionnalités d'un SVI avec les fonctionnalités des SEPT ont un impact positif sur l'engagement des étudiants dans les tâches de rédaction de texte (prise de notes) ;
- Tsaap-Notes est une plateforme perçue comme utile et utilisable par les étudiants.

7.1.2 Contexte

La version initiale de Tsaap-Notes a été testée sur un groupe de quarante étudiants dans le contexte du cours Développement Collaboratif et Logiciels Libres (DCLL). La première partie de l'enseignement est composée de quatre heures de cours en face à face avec la totalité du groupe. Avant le début du premier cours, il a été indiqué aux étudiants que l'utilisation de Tsaap-notes était facultative ; une présentation de l'outil d'une quinzaine de minutes a été faite ; nous avons invité les étudiants à nous donner leur avis au cours de leur utilisation pour améliorer Tsaap-Notes. Toutes les questions qui ont été posées pendant le cours en utilisant le système de votes interactifs de Tsaap-Notes ont été ensuite injectées manuellement dans la plateforme Moodle de l'Université dans l'espace correspondant au cours DCLL.

Table 10. Statistiques sur l'activité de prise de notes

Nombre d'heures de cours	4
Nombre de diapositives présentées	62
Nombre de notes prises par les étudiants	23
Nombre de contributeurs	10 (25%)
Nombre moyen de notes par étudiant	0,58
Nombre moyen de notes par contributeur	2,3
Nombre moyen de notes par diapositives	0,37

Table 11. Impact des votes interactifs

Nombre moyen d'étudiants répondant aux questions	32 (78%)
Nombre de notes prises sur les diapositives faisant l'objet d'au moins une question interactive	15 (65%)
Nombre moyen de notes par diapositive faisant l'objet d'au moins une question interactive	3
Nombre moyen de notes par diapositive ne faisant pas l'objet de question interactive	0,14
Nombre de diapositives faisant l'objet d'au moins une question interactive.	5 (8%)

7.1.3 Résultats quantitatifs

Le Tableau 10 présente les statistiques portant sur l'activité de prise de notes pendant les cours, tandis que le Tableau 11 présente les indicateurs qui permettent d'évaluer l'impact des votes interactifs sur l'engagement des étudiants. Le premier résultat notable est le pourcentage d'étudiants ayant répondu aux questions : 78% soit 32 étudiants. Ce nombre confirme l'efficacité du système de votes interactifs sur la capacité à engager le plus grand nombre d'étudiants. Dix étudiants ont participé à la prise de notes collaborative. Il est remarquable que le nombre de notes prises sur des diapositives faisant l'objet d'au moins une question interactive représente 65% de toutes les notes prises. Enfin, le nombre de diapositives proposant des questions interactives ne représente que 8% de toutes les diapositives présentées pendant les cours. En résumé, 65% de la prise de notes collaborative est réalisée sur 8% des diapositives, celles associées à des questions interactives.

7.1.4 Résultats qualitatifs

Nous avons demandé aux étudiants de remplir un questionnaire afin d'obtenir un retour qualitatif sur Tsaap-Notes. Nous avons obtenu 25 réponses. Le Tableau 12

Table 12. Résultats obtenus sur les questions de type "Oui/Non"

	Oui	Non
Prévoyez-vous d'utiliser Tsaap-Notes pour réviser votre cours avant l'examen ?	22 (88%)	3 (12%)
Prévoyez vous de rejouer les questions interactives dans Moodle ?	7 (28%)	18 (72%)

présente les résultats obtenus. 22 étudiants (sur 25) prévoient d'utiliser Tsaap-Notes lors de la préparation de leurs examens et 7 prévoient de rejouer les questions dans Moodle. Ces résultats nous semblent confirmer la pertinence de Tsaap-notes pour les étudiants.

Enfin, nous avons demandé aux étudiants de qualifier leur expérience avec Tsaap-Notes à l'aide d'une échelle de Likert de cinq niveaux : 1 pour "expérience très négative», 5 pour «expérience très positive". La valeur moyenne obtenue est de 3,68. Les étudiants ont globalement perçu l'utilisation de Tsaap-Notes comme une expérience positive.

7.1.5 Synthèse des résultats

Cette première expérience sur l'utilisation de Tsaap-Notes (1) confirme les résultats des travaux sur les systèmes de votes interactifs : les questions interactives proposées durant le cours engagent la grande majorité des étudiants, la plupart d'entre eux répondant aux questions, (2) montre que le retour immédiat apporté par la présentation des résultats des votes interactifs renforce la prise de notes collaborative : la majorité des notes ont été prises après la présentation des résultats d'une question interactive.

Malgré ce dernier constat positif, le taux de participation à l'activité de prise de notes reste très en dessous du taux de participation aux activités de votes interactifs. Ce faible taux de participation démontre que le maintien d'un taux de participation élevé sur des activités exigeant de la saisie en texte libre est un verrou que la première version de Tsaap-Notes n'a pas su lever.

7.2 Expérimentation suite à l'itération 2

7.2.1 Objectifs

À l'issue de la deuxième itération, Tsaap-Notes propose le dispositif *Notes as Feedback* permettant de générer de manière semi-automatique des tests de révision recyclant les questions et les notes issues de l'utilisation de Tsaap-Notes durant les cours en face à face. Cette fonctionnalité a été introduite dans le but de faciliter l'utilisation de tests sommatifs à visée formative. L'expérimentation menée à l'issue de la deuxième itération avait pour objectif de valider les hypothèses suivantes :

- le recyclage des notes des étudiants pour la constitution de feedback a un impact positif sur l'engagement des étudiants dans des tâches de rédaction de texte (prise de notes) ;
- les étudiants participent massivement aux tests de révisions ;
- les étudiants réussissent mieux aux évaluations sommatives quand celles-ci contiennent des questions posées en cours via Tsaap-Notes et ré-utilisées dans les tests de révision.

7.2.2 Contexte

Le dispositif *Notes as Feedback* a été expérimenté 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 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 par le biais de la plateforme 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%.

7.2.3 Participation à la prise de notes sur les questions interactives

Au démarrage de l'expérimentation (durant la première heure de cours), la plateforme 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 plateforme, soit un taux de 10%. Au démarrage de la 7ème 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%.

Le premier résultat notable de cette expérimentation est que la réutilisation des notes prises par les étudiants dans les tests de révision a sensiblement renforcé l'activité de prise de notes sur les questions interactives : l'utilité perçue des notes a incité les étudiants à la prise de notes.

7.2.4 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 feedbacks dès leur mise à disposition. 47 étudiants, soit 87% des inscrits, ont réalisé les tests de révision avec feedback.

Apparaît ici un deuxième résultat notable : les étudiants préfèrent à l'unanimité jouer les tests avec feedback 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 une aide durant les phases d'auto-évaluation informatisée.

Notons que notre expérimentation n'a pas considéré les tests avec feedback fourni par l'enseignant. Nous avons délibérément écarté cette option car notre objectif est (1) d'automatiser autant que possible la génération du feedback pour chaque question, (2) d'inciter les étudiants à prendre davantage de notes pendant le cours, et (3) de renforcer l'utilisation des étudiants comme ressources d'instruction pour les

autres. Fournir des tests avec feedback fourni par l'enseignant implique d'une part que l'enseignant ait saisi (manuellement) l'ensemble du feedback, et d'autre part que les notes prises par les étudiants ne soient plus valorisées car remplacées par la saisie de l'enseignant. Ces deux implications rentrent directement en contradiction avec nos objectifs initiaux.

7.2.5 Résultats au test d'évaluation sommative

À la fin des périodes 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 était en place depuis deux ans uniquement) : 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 doit cependant être considérée avec précaution 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, qui intègre la plate-forme Tsaap-Notes ainsi qu'un ensemble de tests d'auto-évaluation semi-automatiquement générés et fournissant un feedback issu des notes prises par les étudiants, a entraîné une hausse remarquable des résultats des étudiants sur la partie d'examen informatisée. Il est cependant impossible de discerner ce qui, dans ce dispositif, a eu l'influence la plus notable : en 2013, la plateforme Tsaap-Notes n'était pas intégrée à l'enseignement expérimenté ici et les tests d'auto-évaluation n'avaient pas été proposés aux étudiants.

7.2.6 Synthèse des résultats

Les résultats de la première expérimentation mettant en œuvre le dispositif *Notes as Feedback* sont encourageants : (1) les étudiants s'impliquent davantage dans la prise de notes quand celles-ci sont réutilisées dans les tests d'auto-évaluation ; (2) les tests contenant les notes des étudiants comme source de feedback sont fortement utilisés par les étudiants ; (3) la mise en place de l'ensemble du dispositif (i.e.

Tsaap-Notes et tests d'auto-évaluation) a un impact très positif sur les résultats obtenus par les étudiants à leur examen informatisé.

7.3 Expérimentation suite à l'itération 3

7.3.1 Objectifs

À l'issue de l'itération 3, Tsaap-Notes propose une implantation du processus en N phases décrit dans la section 4.7. Tsaap-Notes supporte ainsi dans cette 3ème version :

- la saisie de texte libre pour fournir un élément de réponse textuel associé à une question posée via Tsaap-Notes ;
- la mise en situation de conflit socio-cognitif pour susciter explicitement la confrontation de points de vue ;
- l'évaluation par les pairs des réponses textuelles.

La troisième expérimentation avait pour objectif la validation des hypothèses suivantes :

- le fait de proposer une évaluation en temps réels des réponses textuelles engage fortement les étudiants dans les tâches de rédaction de texte (rédaction des explications) ;
- la mise en situation de conflit socio-cognitif a un impact positif sur la quantité de bonnes réponses fournies à la deuxième soumission ;
- l'évaluation par les pairs suscite un fort engagement des étudiants et est valide vis-à-vis de l'usage qui en est fait (constitution du feedback dans les tests de révision) ;
- les étudiants perçoivent Tsaap-Notes comme ayant une bonne utilisabilité.

7.3.2 Contexte

La plateforme Tsaap-Notes intégrant le dispositif de soumission de réponse en plusieurs phases (dispositif N-Phases dans la suite du document) a été testé avec deux groupes d'étudiants dans le cadre de deux cours de Master 1 mention Informatique. L'expérimentation s'est déroulée sur 3 sessions de 2 heures pendant lesquelles 5 questions mobilisant le dispositif N-phase ont été posées aux étudiants.

Table 13. Taux de participation aux réponses soumises en N phases

Code cours	Code question	Nombre présents	Nombre réponses soumises	Taux de participation
DCLL	DCLL-1	51	37	72.5%
DCLL	DCLL-2	51	41	80.4%
MA	MA-1	34	34	100%
MA	MA-2	34	34	100%
MA	MA-3	34	32	94.1%
Moyenne	-	40.8	35.6	87.8%

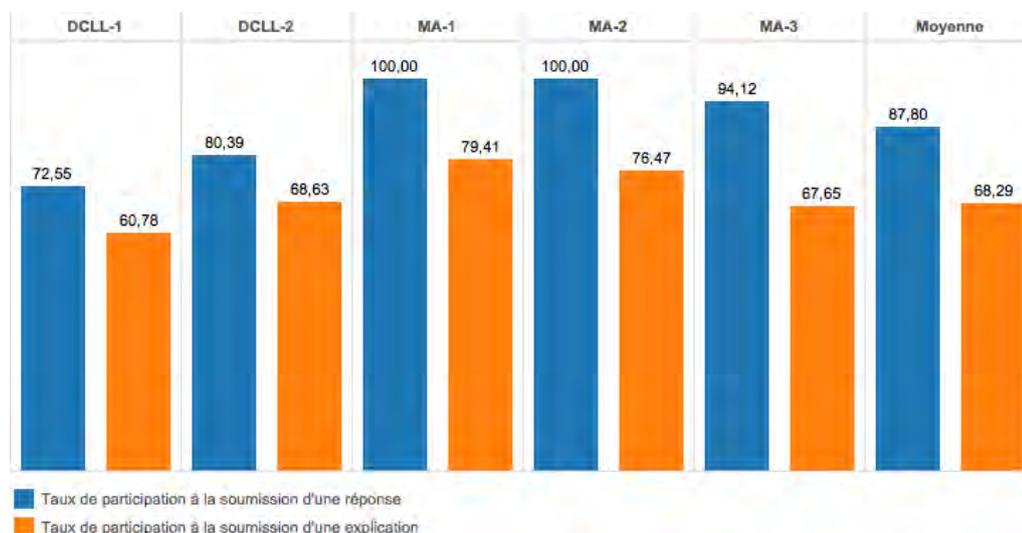


Figure 33. Taux de participation sur les soumissions de réponses et sur les soumissions d'explications

7.3.3 Résultats quantitatifs

Le Tableau 13 montre que 88% des étudiants ont soumis une réponse à chaque question ; ce taux de participation élevé est cohérent avec les autres études relatives aux systèmes de votes interactifs.

Un indicateur remarquable est le taux de participation des étudiants dans la soumission d'une explication : la Figure 33 montre que 68,29% des étudiants qui ont répondu à une question l'ont fait en soumettant une explication pour justifier leur réponse. Même si le taux de participation est en dessous de celui observé sur la soumission d'une réponse (87,8%), il représente une très importante augmentation du taux de participation à une activité exigeant de la saisie de texte libre (au mieux 25% constaté à l'issue de la deuxième itération).

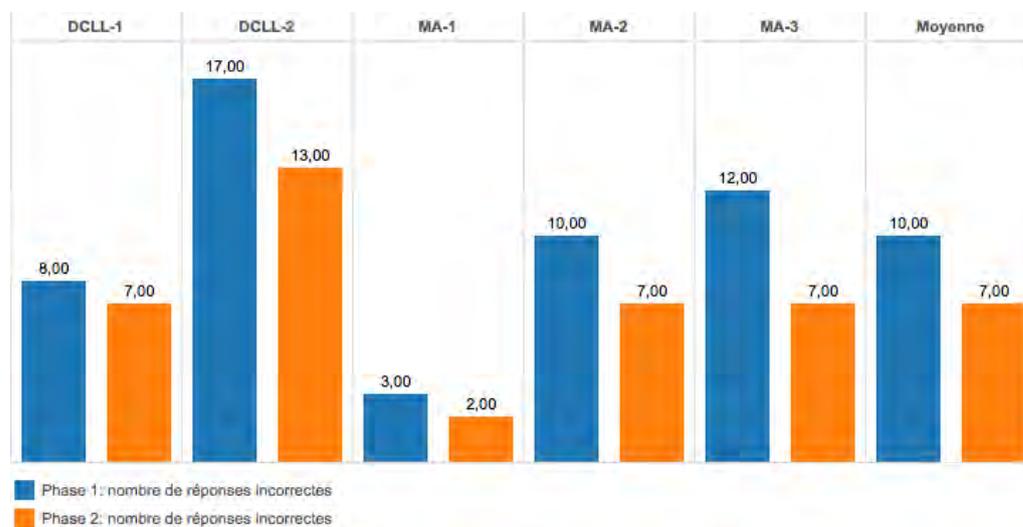


Figure 34. Nombre de mauvaises réponses en phase 1 et en phase 2

La Figure 34 révèle d'autre part qu'entre la première phase et la deuxième, en moyenne 30% des étudiants ayant soumis une mauvaise réponse en première intention ont modifié leur réponse et leur explication. Ceux-ci ont donc pleinement bénéficié du conflit socio-cognitif les ayant amenés à modifier leur croyance sur ce qui était vrai ou faux. Ce résultat est similaire aux résultats obtenus par Crouch et Mazur (2001) concernant les études menées sur l'approche "Instruction par les pairs".

Enfin, la Figure 35 indique que 75,6% des étudiants ont participé à la tâche d'évaluation de trois explications fournies par les pairs. Cette tâche est proposée en phase 3 du processus et n'est pas obligatoire. Le nombre moyen d'évaluateurs par explication est de 4,17 pour une moyenne de 40,8 étudiants présents et de 22,8 explications à évaluer. L'écart-type moyen des notes distribuées par les étudiants (comprises entre 1 et 5) sur les explications est de 0.98. Si cet écart-type moyen est non négligeable puisqu'il signifie que les notes d'un étudiant à l'autre varient de 20% en moyenne, il nous semble raisonnable pour l'usage que nous en faisons, à savoir la sélection des explications les mieux notées pour générer le feedback dans les tests de révision.

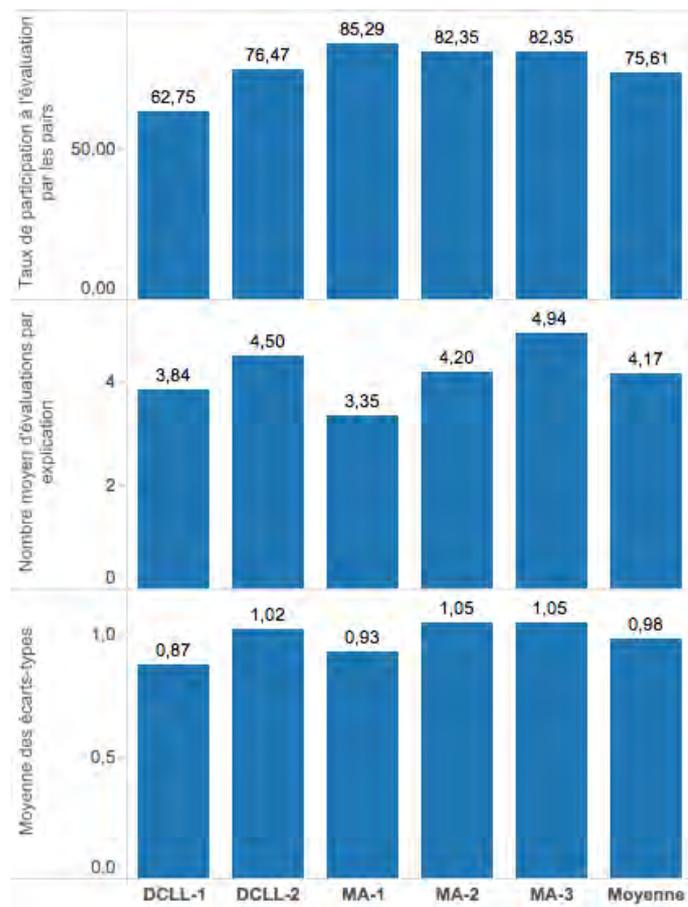


Figure 35. Taux de participation à la tâche d'évaluation par les pairs

Table 14. Les items du *System usability Scale*

	Version originale	Version française
1	I think that I would like to use this system frequently.	Je pense que je voudrais utiliser ce système fréquemment.
2	I found the system unnecessarily complex.	J'ai trouvé le système inutilement complexe.
3	I thought the system was easy to use.	J'ai pensé que le système était facile à utiliser.
4	I think that I would need the support of a technical person to be able to use this system.	Je pense que j'aurais besoin du support d'une personne technique pour être capable d'utiliser le système.
5	I found the various functions in this system were well integrated.	J'ai trouvé que les différentes fonctions du système étaient bien intégrées.
6	I thought there was too much inconsistency in this system.	J'ai pensé qu'il y avait trop de d'incohérence dans ce système.
7	I would imagine that most people would learn to use this system very quickly.	J'imagine que la plupart des gens apprendraient à utiliser le système très rapidement.
8	I found the system very cumbersome to use.	J'ai trouvé le système très lourd à utiliser.
9	I felt very confident using the system.	Je me suis senti à l'aise en utilisant le système.
10	I needed to learn a lot of things before I could get going with this system.	J'ai eu besoin d'apprendre beaucoup de choses avant de pouvoir utiliser le système.

7.3.4 Résultats qualitatifs

Tous les étudiants ayant participé à l'expérimentation ont été contactés pour répondre au questionnaire *System usability Scale*¹ (SUS) (Brooke, 1996). Le SUS s'est imposé comme un standard de l'industrie pour évaluer de manière simple et efficace l'utilisabilité d'un système. Le SUS est composé de dix items dont l'évaluation attendue se répartit sur une échelle de Lickert à cinq niveaux : le premier niveau correspond à l'évaluation "pas du tout d'accord" et le cinquième niveau correspond à l'évaluation "Tout à fait d'accord". Le Tableau 14 présente les dix items du SUS dans leur version originale en anglais et dans leur version traduite en français (version proposée aux étudiants du Master).

Pour calculer le score SUS, il faut commencer par additionner le score de chaque

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

item. Pour les items 1,3,5,7 et 9, le score est le niveau de l'échelle de Lickert moins 1. Pour les items 2,4,6,8 et 10, le score est 5 moins le niveau de l'échelle de Lickert. Chaque score d'item varie de 0 à 4, la somme des scores des 10 items donne un résultat sur 40. Il faut donc multiplier le résultat obtenu par 2,5 pour obtenir le score SUS sur 100. Les recherches sur l'utilisation du SUS (Brooke, 2013) ont conduit à considérer un système comme étant en dessous de la moyenne quand il obtient un score en dessous de 68.

Le SUS permet de mesurer l'utilisabilité perçue d'un système avec un nombre restreint de participants au sondage : entre 8 et 12 (Tullis et Stetson, 2004). Nous avons recueilli 43 réponses au questionnaire ce qui est donc tout à fait recevable pour la mesure de l'utilisabilité de Tsaap-Notes.

Le score SUS obtenu pour Tsaap est de 77,8. Il peut donc être considéré comme un système perçu comme ayant une *bonne* utilisabilité par les étudiants (Bangor *et al.*, 2009). Nous n'avons pas pu proposer le questionnaire aux enseignants car actuellement, en dehors du concepteur de la plateforme, seulement un enseignant utilise Tsaap.

7.3.5 Synthèse des résultats

Les résultats issus de la troisième expérimentation montrent que les étudiants s'engagent largement dans les tâches de rédaction d'explications. La mise en situation de conflit socio-cognitif a un impact positif sur la quantité de réponses correctes fournies après la deuxième soumission. L'évaluation par les pairs engage la grande majorité des étudiants et la qualité des évaluations permet leur utilisation pour identifier et sélectionner les meilleures explications à recycler dans les tests de révision. Enfin, les réponses apportées par les étudiants au questionnaire SUS montrent que le système Tsaap-Notes est perçu comme ayant une bonne utilisabilité.

7.4 Conclusion

Les expérimentations présentées dans ce chapitre avaient un objectif commun : mesurer l'engagement des étudiants dans les différentes activités supportées par Tsaap-Notes. Les résultats de la troisième expérimentation s'intéressant aux impacts de l'implantation du processus en N phases montrent des résultats très encourageants : le système Tsaap-Notes se place dans les systèmes engageant plus

des deux tiers des étudiants présents en cours (critère E-TE), y compris sur les tâches exigeant la fourniture de contributions textuelles.

Le tableau 15 présente l'évaluation de la plateforme Tsaap-Notes pour l'ensemble des critères de la grille d'évaluation présentée en section 3.2.2. Le système que nous avons conçu, en combinant fonctionnalité des SVI et des SEPT et en intégrant un nouveau processus d'évaluation d'audience, parvient à supporter l'ensemble des activités de référence et des stratégies caractéristiques de l'évaluation formative tout en engageant la grande majorité des étudiants.

Table 15. 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														

Conclusion et perspectives de recherche

Chapitre 8

Conclusion et perspectives de recherche

Sommaire

8.1	Les contributions	135
8.1.1	La définition du concept d'évaluation d'audience	136
8.1.2	La modélisation standardisée de processus d'évaluation d'audience	136
8.1.3	Le processus d'évaluation d'audience en N phases	136
8.1.4	La conception et l'implantation du système Tsaap-Notes .	137
8.1.5	Synthèse	137
8.2	Perspectives de recherche	138
8.2.1	Axe interaction homme-machine	138
8.2.2	Axe architecture logicielle	138
8.2.3	Axe jeux sérieux	138
8.2.4	Axe intelligence artificielle	139

8.1 Les contributions

La description détaillée de la plateforme Tsaap-Notes présentée dans la partie II et les résultats des premières expérimentations menées permettent de dresser un bilan des contributions apportées par notre travail.

8.1.1 La définition du concept d'évaluation d'audience

La définition du concept d'évaluation d'audience, centrée sur l'évaluation formative appliquée simultanément à un grand nombre d'étudiants, offre un cadre de travail pour la conception de toute une famille de systèmes permettant la mise en œuvre de processus d'évaluation formative dans les amphithéâtres d'universités. Les systèmes de votes interactifs, pionniers des systèmes d'évaluations d'audience, ouvrent la voie pour la construction de systèmes plus évolués.

8.1.2 La modélisation standardisée de processus d'évaluation d'audience

Différents travaux ont décrit les processus mis en jeu dans l'évaluation formative (Black et Wiliam, 2009) ou dans les systèmes de votes interactifs (Roschelle *et al.*, 2004). Cependant ces descriptions ne s'appuient pas sur des standards de description de processus et ne facilitent donc ni le partage, ni la communication, ni la compréhension de ces processus par l'ensemble des parties prenantes impliquées dans la mise en œuvre de systèmes d'évaluation d'audience.

Nous avons proposé l'utilisation du standard BPMN 2.0 (OMG, 2011) pour la description des processus abordés dans nos travaux. Ainsi, les processus principaux mis en jeu dans les systèmes de votes interactifs et dans la plateforme Tsaap-Notes ont pu être décrits en utilisant une notation standard permettant des discussions et des échanges sur un langage commun.

8.1.3 Le processus d'évaluation d'audience en N phases

Le processus en N phases que nous avons introduit intègre davantage d'activités de référence de l'évaluation formative que les processus traditionnels :

- l'évaluation par les pairs (A4) ;
- l'utilisation de tests sommatifs en vue d'usage formatif (A5).

De plus, ce nouveau processus renforce fortement certaines des stratégies caractérisant l'évaluation formative :

- de nouvelles situations et tâches d'apprentissage sont proposées aux étudiants pour éliciter des preuves de leur niveau d'avancement (S2) ;
- de nouveaux formats de feedback sont proposés aux étudiants pour les faire progresser dans leur apprentissage (S3) ;

- les apprenants produisent des ressources exploitables par les autres (S4).

8.1.4 La conception et l'implantation du système Tsaap-Notes

La conception du système d'évaluation d'audience Tsaap-Notes constitue une autre contribution de notre travail. En combinant les fonctionnalités des systèmes de votes interactifs et des systèmes d'échange et de partage de textes pour supporter le processus en N phases, Tsaap-Notes intègre de nouvelles formes d'interaction permettant la fourniture d'éléments de réponses saisis sous forme de texte libre et la fourniture de feedbacks textuels. L'évaluation par les pairs supportée par Tsaap-Notes résout la question de l'évaluation simultanée d'un grand nombre de contributions textuelles en confiant en parallèle et de manière "décentralisée" l'évaluation des contributions à un grand nombre d'évaluateurs (les étudiants). En supportant la mise en oeuvre d'une situation de conflit socio-cognitif pour chaque étudiant, Tsaap-Notes suscite explicitement la confrontation d'idées entre pairs. Enfin, la plateforme fournit un dispositif d'export des questions et de contributions qui facilite la tâche des enseignants souhaitant s'inscrire dans la stratégie d'utilisation des tests sommatifs pour des usages formatifs.

8.1.5 Synthèse

En s'inscrivant dans le cadre de l'évaluation d'audience et du support du processus en N phases, la conception et la mise en oeuvre du système Tsaap-Notes a permis d'apporter des réponses aux deux questions de recherche énoncées en introduction :

1. le système Tsaap-Notes enrichit la mise en oeuvre des évaluations d'audience en introduisant de nouvelles activités de références, de nouvelles interactions, de nouveaux types de feedbacks et en renforçant les stratégies caractérisant l'évaluation formative.
2. les expérimentations ont montré que Tsaap-Notes engageait fortement les étudiants participant aux cours utilisant le système développé, y compris sur des tâches exigeant la fourniture de contributions textuelles.

8.2 Perspectives de recherche

L'introduction des concepts d'évaluation d'audience et de systèmes d'évaluation d'audience offre un cadre de travail dans lequel peuvent s'inscrire de nombreux axes de recherche dans le champ informatique.

8.2.1 Axe interaction homme-machine

La plateforme Tsaap-Notes introduit le support d'activités et de stratégies relatives à l'évaluation formative pratiquée dans les amphithéâtres d'universités. Certaines de ces activités ne sont pas couvertes de manière optimale par Tsaap-Notes. La clarification des objectifs d'apprentissage et des critères de succès n'est pas directement supportée par notre système. Les questions ouvertes ne sont pas non plus complètement adressées par Tsaap-Notes qui propose une soumission de réponse à des questions à choix exigeant une justification. Le support des questions ouvertes dans un système d'évaluation d'audience sous l'angle des interactions homme-machine à mettre en oeuvre constitue un axe de recherche.

La génération de tests de révisions à partir de Tsaap-Notes nécessite encore des tâches manuelles. Nous travaillons actuellement à une plus forte inter-opérabilité de la plateforme avec les différents *Learning Management System* (LMS) du marché en implantant le protocole de communication IMS LTI¹. Cette intégration avec les LMS permettra d'automatiser davantage les tâches facilitant par exemple l'utilisation de tests sommatifs à visée formative.

8.2.2 Axe architecture logicielle

Nous avons introduit l'utilisation de BPMN pour la description des processus. BPMN a été conçu pour décrire des modèles exécutables. Explorer la faisabilité et la pertinence de l'utilisation d'une approche d'ingénierie dirigée par les modèles pour concevoir les systèmes d'évaluation d'audience constitue un axe de recherche.

8.2.3 Axe jeux sérieux

Les jeux sérieux sont aujourd'hui une piste privilégiée pour améliorer la motivation des étudiants durant leur apprentissage. Des recherches s'interrogeant sur

1. <http://www.msglobal.org/toolsinteroperability2.cfm>

la manière d'intégrer des éléments caractéristiques des jeux sérieux pour améliorer la motivation des étudiants pendant les cours en amphithéâtre pourraient être menées.

8.2.4 Axe intelligence artificielle

Le système que nous avons conçu règle la question de l'évaluation des contributions textuelles pendant le cours en mobilisant les étudiants pour pratiquer l'évaluation par les pairs. Des recherches sur l'évaluation de contributions écrites par une intelligence artificielle sont déjà en cours. Ces recherches trouvent leur terrain d'application dans les cours ouverts massifs en ligne. Il serait intéressant d'étendre ces recherches au contexte des évaluations d'audience pour offrir de nouvelles possibilités d'évaluations des contributions fournies sous forme de texte libre.

Dans le cadre de la mise en œuvre du support des questions ouvertes, l'algorithme de génération de conflits socio-cognitifs ne pourra plus s'appuyer sur la classification des explications reposant sur le fait qu'elles correspondent ou non à la bonne réponse. De nouveaux algorithmes capables de générer un conflit socio-cognitif devront alors être étudiés.

Enfin, s'appuyant sur les avancées du champ du *Learning Analytics*, la génération de tests de révisions pourrait faire l'objet de recherche afin d'offrir de véritables tests personnalisés à chaque étudiant, tant au niveau des questions que des feedbacks textuels proposés. Nous mettons actuellement en place différentes expérimentations dans les contextes de l'enseignement secondaire et universitaire afin de collecter le plus grand nombre de données possible et d'explorer concrètement cet axe de recherche.

Bibliographie

- Abrahamson, A. (1999). Teaching with classroom communication system—what it involves and why it works. *In 7th International Workshop New Trends in Physics Teaching, Puebla, Mexico.*
- Aksoy, L., van Riel, A., Kandampully, J., Bolton, R. N., Parasuraman, A., Hoefnagels, A., Migchels, N., Kabadayi, S., Gruber, T., Komarova Loureiro, Y. *et al.* (2013). Understanding generation y and their use of social media : a review and research agenda. *Journal of Service Management*, 24(3):245–267.
- Aleven, V. (2010). Rule-based cognitive modeling for intelligent tutoring systems. *In Advances in intelligent tutoring systems*, pages 33–62. Springer.
- Aleven, V., McLaren, B. M., Sewall, J. et Koedinger, K. R. (2006). The cognitive tutor authoring tools (ctat) : preliminary evaluation of efficiency gains. *In Intelligent Tutoring Systems*, pages 61–70. Springer.
- Anderson, J. R., Boyle, C. F. et Yost, G. (1985). The geometry tutor. *In IJCAI*, pages 1–7.
- Bangor, A., Kortum, P. et Miller, J. (2009). Determining what individual sus scores mean : Adding an adjective rating scale. *Journal of usability studies*, 4(3):114–123.
- Beatty, I. D. (2005). Transforming student learning with classroom communication systems. *arXiv preprint physics/0508129*.
- Beatty, I. D. et Gerace, W. J. (2009). Technology-enhanced formative assessment : A research-based pedagogy for teaching science with classroom response technology. *Journal of Science Education and Technology*, 18(2):146–162.
- Black, P., Harrison, C. et Lee, C. (2003). *Assessment for learning : Putting it into practice*. McGraw-Hill Education (UK).

- Black, P. et Wiliam, D. (1998). Assessment and classroom learning. *Assessment in education*, 5(1):7–74.
- Black, P. et Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability (formerly : Journal of Personnel Evaluation in Education)*, 21(1):5–31.
- Bloom, B. S. (1971). *Handbook on formative and summative evaluation of student learning*. McGraw-Hill Book Company, 330 West 42nd Street, New York, New York 10036.
- Bloom, B. S. (1984). The 2 sigma problem : The search for methods of group instruction as effective as one-to-one tutoring. *Educational researcher*, pages 4–16.
- Borau, K., Ullrich, C., Feng, J. et Shen, R. (2009). Microblogging for language learning : Using twitter to train communicative and cultural competence. *In Advances in Web Based Learning–ICWL 2009*, pages 78–87. Springer.
- Boud, D. (2013). *Enhancing learning through self-assessment*. Routledge.
- Brewer, C. A. (2004). Near real-time assessment of student learning and understanding in biology courses. *BioScience*, 54(11):1034–1039.
- Brooke, J. (1996). Sus-a quick and dirty usability scale. *Usability evaluation in industry*, 189(194):4–7.
- Brooke, J. (2013). Sus : a retrospective. *Journal of Usability Studies*, 8(2):29–40.
- Bull, J. et McKenna, C. (2003). *A blueprint for computer-assisted assessment*. Routledge.
- Bullock, D., LaBella, V., Clingan, T., Ding, Z., Stewart, G. et Thibado, P. (2002). Enhancing the student-instructor interaction frequency. *The Physics Teacher*, 40(9):535–541.
- Burnstein, R. A. et Lederman, L. M. (2001). Using wireless keypads in lecture classes. *The Physics Teacher*, 39(1):8–11.
- Butler, D. L. et Winne, P. H. (1995). Feedback and self-regulated learning : A theoretical synthesis. *Review of educational research*, 65(3):245–281.

- Butler, R. (1987). Task-involving and ego-involving properties of evaluation : Effects of different feedback conditions on motivational perceptions, interest, and performance. *Journal of educational psychology*, 79(4):474.
- Caldwell, J. E. (2007). Clickers in the large classroom : Current research and best-practice tips. *CBE-Life Sciences Education*, 6(1):9–20.
- Chanock, K. (2000). Comments on essays : do students understand what tutors write? *Teaching in Higher Education*, 5.
- Cheng, A., Evans, M. et Singh, H. (2009). Inside twitter : An in-depth look inside the twitter world. *Report of Sysomos, June, Toronto, Canada*.
- 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. *Mathematics Teacher*, 100(2):100–104.
- Crooks, T. J. (1988). The impact of classroom evaluation practices on students. *Review of educational research*, 58(4):438–481.
- Crouch, C. H. et Mazur, E. (2001). Peer instruction : Ten years of experience and results. *American journal of physics*, 69(9):970–977.
- Cutts, Q. (2006). Practical lessons from four years of using an ars in every lecture of a large class. *Audience response systems in higher education*, pages 65–79.
- Da Costa, J. et Schneider, D. K. (2015). Modélisation et implémentation de dispositifs pédagogiques avec bpmn 2.0. *In Actes de la 7ème conférence sur les Environnements Informatiques pour l'Apprentissage Humain - EIAH 2015*.
- Delangue, C. (2012). Interview clément delangue, fondateur unishared. <http://www.le-toaster.fr/interweb/interview-avec-clement-delangue-fondateur-dunishared/>. Accessed : 2015-08-04.
- Devenney, R. (1989). How esl teachers and peers evaluate and respond to student writing. *RELC Journal : A Journal of Language Teaching and Research in Southeast Asia*, 20(1):77–90.

- Dochy, F., Segers, M. et Sluijsmans, D. (1999). The use of self-, peer and co-assessment in higher education : A review. *Studies in Higher education*, 24(3):331–350.
- Dodero, J. M., Tattersall, C., Burgos, D. et Koper, R. (2008). Transformational techniques for model-driven authoring of learning designs. *In Advances in Web Based Learning–ICWL 2007*, pages 230–241. Springer.
- Draper, S. W. et Brown, M. I. (2004). Increasing interactivity in lectures using an electronic voting system. *Journal of computer assisted learning*, 20(2):81–94.
- Dufresne, R. J. et Gerace, W. J. (2004). Assessing-to-learn : Formative assessment in physics instruction. *The Physics Teacher*, 42(7):428–433.
- Dufresne, R. J., Gerace, W. J., Leonard, W. J., Mestre, J. P. et Wenk, L. (1996). Classtalk : A classroom communication system for active learning. *Journal of computing in higher education*, 7(2):3–47.
- Dwyer, C. A. (2008). *The future of assessment : shaping teaching and learning*. Routledge.
- Eberly Center, C. M. U. (2015). What is the difference between formative and summative assessment ? <https://www.cmu.edu/teaching/assessment/basics/formative-summative.html>. Accessed : 2015-07-27.
- Ebner, M. (2009). Interactive lecturing by integrating mobile devices and micro-blogging in higher education. *CIT. Journal of Computing and Information Technology*, 17(4):371–381.
- Ebner, M., Lienhardt, C., Rohs, M. et Meyer, I. (2010). Microblogs in higher education—a chance to facilitate informal and process-oriented learning? *Computers & Education*, 55(1):92–100.
- Elliott, C. (2003). Using a personal response system in economics teaching. *International Review of Economics Education*, 1(1):80–86.
- Falchikov, N. (1995). Peer feedback marking : developing peer assessment. *Programmed Learning*, 32(2):175–187.

- Fies, C. et Marshall, J. (2006). Classroom response systems : A review of the literature. *Journal of Science Education and Technology*, 15(1):101–109.
- Fredricks, J. A., Blumenfeld, P. C. et Paris, A. H. (2004). School engagement : Potential of the concept, state of the evidence. *Review of educational research*, 74(1):59–109.
- Gauci, S. A., Dantas, A. M., Williams, D. A. et Kemm, R. E. (2009). Promoting student-centered active learning in lectures with a personal response system. *Advances in Physiology Education*, 33(1):60–71.
- Greer, L. et Heaney, P. J. (2004). Real-time analysis of student comprehension : An assessment of electronic student response technology in an introductory earth science course. *Journal of Geoscience Education*, 52:345–351.
- Grosseck, G. et Holotescu, C. (2008). Can we use twitter for educational activities. *In 4th international scientific conference, eLearning and software for education, Bucharest, Romania.*
- Habert, B. (2012). Notes de cours collaboratives. <http://archinfo02.hypotheses.org/38>. Accessed : 2015-08-04.
- Hattie, J. et Jaeger, R. (1998). Assessment and classroom learning : A deductive approach. *Assessment in Education*, 5(1):111–122.
- Hattie, J. et Timperley, H. (2007). The power of feedback. *Review of educational research*, 77(1):81–112.
- Higgins, R. (2000). *Be more critical : rethinking assessment feedback*. DYE. V, PC4000.
- Higgins, R., Hartley, P. et Skelton, A. (2002). The conscientious consumer : reconsidering the role of assessment feedback in student learning. *Studies in Higher Education*, 27(1):53–64.
- IMS, L. (2003a). Ims learning design best practice and implementation guide. http://www.imsglobal.org/learningdesign/ldv1p0/imsld_bestv1p0.html#1503669. Accessed : 2015-08-10.

- IMS, L. (2003b). Learning design specification v1. <http://www.imsglobal.org/learningdesign/>. Accessed : 2015-08-10.
- IMS, Q. (2012). Ims question & test interoperability specification. <http://www.imsglobal.org/question/>.
- Ivanic, R., Clark, R. et Rimmershaw, R. (2000). *What am I supposed to make of this? : the messages conveyed to students by tutors' written comments*. Open University Press.
- Kam, M., Wang, J., Iles, A., Tse, E., Chiu, J., Glaser, D., Tarshish, O. et Canny, J. (2005). Livenotes : a system for cooperative and augmented note-taking in lectures. *In Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 531–540. ACM.
- Kay, R. H. et LeSage, A. (2009). Examining the benefits and challenges of using audience response systems : A review of the literature. *Computers & Education*, 53(3):819–827.
- Kennedy, G. E. et Cutts, Q. I. (2005). The association between students' use of an electronic voting system and their learning outcomes. *Journal of Computer Assisted Learning*, 21(4):260–268.
- Kiewra, K. A., DuBois, N. F., Christian, D. et McShane, A. (1988). Providing study notes : Comparison of three types of notes for review. *Journal of Educational Psychology*, 80(4):595.
- Kluger, A. N. et DeNisi, A. (1996). The effects of feedback interventions on performance : a historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological bulletin*, 119(2):254.
- Koh, Y. (2014). Report : 44% of twitter accounts have never sent a tweet. <http://blogs.wsj.com/digits/2014/04/11/new-data-quantifies-dearth-of-tweeters-on-twitter/>. Accessed : 2015-08-05.
- Laforcade, P. (2004). *Méta-modélisation UML pour la conception et la mise en oeuvre de situations-problèmes coopératives*. Thèse de doctorat, Université de Pau et des Pays de l'Amour.

- Lakatos, I., Malamoud, C. et Spitz, J.-F. (1994). *Histoire et méthodologie des sciences : programmes de recherche et reconstruction rationnelle*. Presses universitaires de France.
- Lea, M. (2001). Computer conferencing and assessment : new ways of writing in higher education. *Studies in Higher Education*, 26(2):163–181.
- Liu, M. (1997). *Fondements et pratiques de la recherche-action*. l’Harmattan.
- Martinez-Ortiz, I., Sierra, J.-L. et Fernandez-Manjon, B. (2009). Authoring and reengineering of ims learning design units of learning. *Learning Technologies, IEEE Transactions on*, 2(3):189–202.
- Mazur, E. (1997). Peer instruction : getting students to think in class. *In AIP Conference Proceedings*, pages 981–988. IOP INSTITUTE OF PHYSICS PUBLISHING LTD.
- McLoone, S., Villing, R. et O’Keeffe, S. (2015). A novel smart device student response system for supporting high quality active learning in the engineering and science disciplines. *AISHE-J : The All Ireland Journal of Teaching and Learning in Higher Education*, 7(2).
- Meisels, S. J. (2003). Impact of instructional assessment on elementary children’s achievement. *education policy analysis archives*, 11:9.
- MENESR (2014). Repères et références statistiques enseignements formation recherche. Rapport technique, Ministère de l’Éducation Nationale, de l’Enseignement Supérieur et de la Recherche.
- Miller, T. (2009). Formative computer-based assessment in higher education : the effectiveness of feedback in supporting student learning. *Assessment & Evaluation in Higher Education*, 34(2):181–192.
- Mitrovic, A. (2010). Modeling domains and students with constraint-based modeling. *In Advances in intelligent tutoring systems*, pages 63–80. Springer.
- Nicol, D. J. et 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. et 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.
- Nodenot, T. (2007). Scénarisation pédagogique et modèles conceptuels d'un eiah : Que peuvent apporter les langages visuels? *Revue Internationale des Technologies en Pédagogie Universitaire (RITPU)/International Journal of Technologies in Higher Education (IJTHE)*, 4(2):85–102.
- OASIS, W. (2007). Web services business process execution language version 2.0. <http://docs.oasis-open.org/wsbpel/2.0/OS/wsbpel-v2.0-OS.html>. Accessed : 2015-08-10.
- OMG, B. (2011). Business process model and notation (bpmn) version 2.0. <http://www.omg.org/spec/BPMN/2.0/>. Accessed : 2015-08-10.
- OMG, U. (2000). Unified modeling language. <http://www.omg.org/spec/UML/>. Accessed : 2015-08-10.
- O'Neill, M. E. (2005). Automated use of a wiki for collaborative lecture notes. *In ACM SIGCSE Bulletin*, volume 37, pages 267–271. ACM.
- Paquette, G., Léonard, M., Lundgren-Cayrol, K., Mihaila, S. et Gareau, D. (2006). Learning design based on graphical knowledge-modelling. *Journal of Educational Technology & Society*, 9(1):97–112.
- Perrenoud, P. (1989). *Pour une approche pragmatique de l'évaluation formative*. Conseil de l'Europe.
- Rabardel, P. (1995). *Les hommes et les technologies; approche cognitive des instruments contemporains*. Armand Colin.
- Ramaprasad, A. (1983). On the definition of feedback. *Behavioral Science*, 28(1): 4–13.
- Rawlings, A., Van Rosmalen, P., Koper, R., Rodriguez-Artacho, M. et Lefrere, P. (2002). Survey of educational modelling languages (emls). Rapport technique, CEN/ISSS.

- Reay, N. W., Li, P. et Bao, L. (2008). Testing a new voting machine question methodology. *American Journal of Physics*, 76(2):171–178.
- Reinhardt, W., Ebner, M., Beham, G. et Costa, C. (2009). How people are using twitter during conferences. *Creativity and Innovation Competencies on the Web. Proceedings of the 5th EduMedia*, pages 145–156.
- Ricketts, C. et Wilks, S. (2002). Improving student performance through computer-based assessment : insights from recent research. *Assessment & evaluation in higher education*, 27(5):475–479.
- Rodriguez, M. C. (2004). The role of classroom assessment in student performance on timss. *Applied Measurement in Education*, 17(1):1–24.
- Rogriguez, M. C., Derntl, M. et Botturi, L. (2010). Visual instructional design languages. *Journal of Visual Languages & Computing*, 21(6):311–312.
- Roschelle, J., Penuel, W. R. et Abrahamson, L. (2004). Classroom response and communication systems : Research review and theory. *In Annual Meeting of the American Educational Research Association, San Diego, CA*, pages 1–8.
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional science*, 18(2):119–144.
- Shaffer, D. M. et Collura, M. J. (2009). Evaluating the effectiveness of a personal response system in the classroom. *Teaching of Psychology*, 36(4):273–277.
- Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational researcher*, pages 4–14.
- Simon, B., Davis, K., Griswold, W. G., Kelly, M. et Malani, R. (2008). Noteblogging : taking note taking public. *ACM SIGCSE Bulletin*, 40(1):417–421.
- Singh, G., Denoue, L. et Das, A. (2004). Collaborative note taking. *In Wireless and Mobile Technologies in Education, 2004. Proceedings. The 2nd IEEE International Workshop on*, pages 163–167. IEEE.
- Spence, D. (2011). All a-twitter. *BMJ*, 343.
- Steimle, J., Gurevych, I. et Mühlhäuser, M. (2007). Notetaking in university courses and its implications for elearning systems. *In DeLFI*, volume 5, pages 45–56.

- Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of educational Research*, 68(3):249–276.
- Trees, A. R. et Jackson, M. H. (2007). The learning environment in clicker classrooms : student processes of learning and involvement in large university-level courses using student response systems. *Learning, Media and Technology*, 32(1): 21–40.
- Tullis, T. S. et Stetson, J. N. (2004). A comparison of questionnaires for assessing website usability. *In Usability Professional Association Conference*, pages 1–12.
- Uhari, M., Renko, M. et Soini, H. (2003). Experiences of using an interactive audience response system in lectures. *BMC Medical Education*, 3(1):12.
- UNESCO (2009). Recueil de données mondiales sur l'éducation 2009. Rapport technique, Institut de statistiques de l'UNESCO.
- William, D. (2000). Formative assessment in mathematics part 3 : The learner's role. *Equals : Mathematics and Special Educational Needs*, 6(1):19–22.
- William, D. (2007). Keeping learning on track. *Second handbook of research on mathematics teaching and learning*, pages 1053–1098.
- YAMAMOTO, T., OKUNUKI, M., HWANG, W.-Y. et KOBAYASHI, K. (2014). An interactive tool to increase the value of learning. *ICCE 2014*, page 43.
- Yorke, M. (2003). Formative assessment in higher education : Moves towards theory and the enhancement of pedagogic practice. *Higher education*, 45(4):477–501.
- Zakrzewski, S. et Bull, J. (1998). The mass implementation and evaluation of computer-based assessments. *Assessment & Evaluation in Higher Education*, 23(2):141–152.

Résumé

La croissance importante du nombre d'étudiants et le changement de profils des apprenants constatés durant les deux dernières décennies ont bouleversé l'enseignement supérieur. Le besoin d'adapter les enseignements a émergé brutalement. Les cours en face à face s'adressant à un grand nombre d'étudiants offrent un cadre soutenable pour la recherche et la mise en œuvre de nouvelles pratiques pédagogiques. Dans ce contexte, différentes études montrent que l'utilisation de systèmes de votes interactifs (SVI) est un moyen de proposer des évaluations formatives aux étudiants durant le cours en les engageant massivement. Cependant ces systèmes n'intègrent pas toutes les dimensions caractérisant l'évaluation formative, ils ne tirent pas parti des apports des systèmes d'échanges et de partage de textes (SEPT) ayant fait leur apparition avec Internet et offrent donc des modalités d'interactions limitées. Ainsi nos travaux adressent les deux questions de recherche suivantes :

- comment enrichir la mise en œuvre d'évaluations formatives dans le contexte d'un cours en face à face dispensé à un grand nombre d'étudiants ?
- comment susciter l'engagement du plus grand nombre d'étudiants dans les activités relatives à ces évaluations formatives ?

Pour répondre à ces questions, nous proposons en premier lieu un cadre de travail pour la conception de systèmes d'évaluation formative dans le contexte des cours en face à face adressés à une large audience. Nous introduisons ainsi le concept d'évaluation d'audience et décrivons les processus qui lui sont associés en utilisant un langage standard de modélisation. De plus, afin d'enrichir la mise en œuvre de l'évaluation formative dans les amphithéâtres, nous proposons un nouveau processus baptisé "processus en N phases" intégrant des activités et stratégies caractéristiques de l'évaluation formative telles que l'évaluation par les pairs ou l'utilisation des étudiants comme ressources pour les autres. Ce processus supporte également des modalités d'interaction plus riches qui incluent la saisie de texte libre ou la fourniture de feedback sous forme textuelle. La deuxième partie de nos travaux porte sur la conception et la mise en œuvre du système d'évaluation d'audience Tsaap-Notes qui combine les atouts des SVI et des SEPT pour implanter le processus en N phases. Ce système d'évaluation d'audience enrichi a fait l'objet d'une série d'expérimentations montrant des résultats probant sur la capacité du système à maintenir l'engagement de la très grande majorité des étudiants dans le processus d'évaluation formative.

Abstract

The significant growth in student numbers and changing learner profiles observed during the last two decades have shaken higher education. The need to adapt teaching strongly appeared. Face to face lectures for a large audience are a sustainable frame for research and implementation of new pedagogical practices. In this context, different studies show that the use of audience response system (ARS) is an efficient way to provide students with formative assessments during lectures and that they engage a large numbers of attendees. However, these systems do not integrate all dimensions of formative assessments, they do not take advantages of sharing text systems (STS) who appeared with Internet and therefore provide only limited types of interaction. Thus, our work addresses the following research questions :

- how to enhance implementation of formative assessments in the context of face to face lectures adressed to a large number of students ?
- how to maintain engagement of the majority of students in activities related to these formative assessments ?

In order to answer these questions, we first propose a framework for the design of formative assessment systems to be used during face to face lectures addressed to a large audience. Thus, we introduce the concept of Audience Assessment and describe its related processes using a standard language. Then, in order to enhance implementation of formative assessments addressed to a large audience, we propose a new process called "N phases process" which introduces characteristic activities and strategies of formative assessments such as peer assessment or use of students as resources of instruction for others. This process supports also richer types of interaction such as answer given as free text entry or feedback provided as text.

The second part of our work focuses on the design and the implementation of the audience assessment system called « Tsaap-Notes ». Tsaap-Notes combines the strengths of SVI and STS to implement the N phases process. A series of experimentations has been conducted on Tsaap-Notes showing good results on the capacity of the N phases process to engage the large majority of students.

