

Suivi des apprentissages au moyen d'évaluations formatives par questions à choix multiples diffusées sur le Web par le logiciel eTests

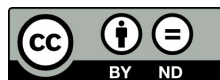
Compte rendu d'expérience intégrant les TIC

Résumé

eTests est un logiciel libre permettant l'édition collaborative et la diffusion d'évaluations formatives par questions à choix multiples sur le Web. Ces évaluations sont formatives, car à chaque proposition de réponse sont associés des justificatifs détaillés ou des indices, selon que la proposition est correcte ou incorrecte. Ce retour instantané d'information permet à l'étudiant de rectifier une éventuelle erreur ou de conforter son raisonnement. L'enregistrement de chaque réponse permet le suivi des performances individuelles ou de groupes. Ceci permet à l'apprenant de se positionner vis-à-vis de ses pairs et le responsabilise dans sa formation. L'analyse éduométrique des réponses permet de détecter des difficultés de résolution, qui peuvent révéler la présence d'un obstacle cognitif de la matière enseignée, dont le diagnostic pourra être affiné par la rédaction de séquences de questions de complexité croissante.

Mots-clés

Évaluation formative, QCM, logiciel libre, *open source*, obstacle cognitif, biostatistique



©Auteur(s). Cette œuvre, disponible à http://ritpu.ca/IMG/pdf/RITPU_v11_n01_19.pdf, est mise à disposition selon les termes de la licence Creative Commons Attribution - Pas de Modification 2.5 Canada : <http://creativecommons.org/licences/by-nd/2.5/ca/deed.fr>

Grégoire **VINCKE**

Université de Namur, Belgique, gregoire.vincke@uclouvain.be

Benoît **BIHIN**

Université de Namur, Belgique, benoit.bihin@unamur.be

Anne-Cécile **WAUTHY**

Université de Namur, Belgique, anne-cecile.wauthy@unamur.be

Eliane **AI ZIND**

Université de Namur, Belgique, elianezind@hotmail.com

Arnaud **VERVOORT**

Université de Namur, Belgique, arnaud.vervoort@unamur.be

Eric **DEPIEREUX**

Université de Namur, Belgique, eric.depiereux@unamur.be

Abstract

eTests is a free software that can be used to offer web-based multiple-choice formative assessment questionnaires. The assessments are formative because each choice is linked with complementary information in the form of hints if the answer is incorrect or comments if the answer is correct. Instant feedback allows students to correct any errors or confirm their reasoning. The answers are recorded so that individual or group performance can be monitored. This transfers responsibility to the learners by helping them compare themselves with their peers. The teachers can perform an edumetric analysis of the distribution of the answers to detect resolution difficulties, which could be the sign of a cognitive obstacle in the learning process. Creating increasingly complex sequences of questions can refine the diagnosis of any such obstacles.

Keywords

Formative assessment, freeware, open source, cognitive obstacle, biostatistics

Introduction

Objectif

Cette étude s'insère dans un processus de délocalisation numérique, initié par la création d'une plateforme d'autoapprentissage (Vincke et Depiereux, 2010), d'une partie d'un dispositif d'apprentissage de la biostatistique. Le public cible est constitué d'étudiants inscrits en seconde année de premier cycle en sciences biologiques et médicales à l'Université de Namur (Belgique). Cette hybridation du dispositif d'apprentissage a permis la mutation du contrat didactique en recentrant le temps présentiel sur la résolution d'obstacles cognitifs (Vincke, Wauthy, Bihin et Depiereux, 2013). Cet article se focalise sur une innovation technologique, la plateforme eTests, en situant son utilisation didactique par rapport aux principaux modèles d'enseignement et d'apprentissage et en analysant les perspectives qu'elle offre en matière d'évaluation formative et de détection d'obstacles cognitifs.

Évolution du dispositif d'apprentissage et d'évaluation

Notre premier dispositif d'apprentissage sur ordinateur (Van Vyve-Genette, Gohy et Feytmans, 1988) fut relayé à partir des années 2000 par un site Web statique (Calmant, 2004) puis dynamique (Vincke et Depiereux, 2010).

Depuis 1990, en réponse au passage rapide d'une université de classe à une université de masse (Albero, 2011), l'évaluation certificative est réalisée au moyen de QCM avec coefficient de certitude, selon une méthodologie inspirée de Leclercq (1982). En 2002, certains enseignants ont souhaité utiliser à des fins formatives certaines des questions certificatives, considérées comme confidentielles, en les rendant accessibles sur le Web. La tension qui émergea entre les partisans du rôle formatif ou certificatif d'une évaluation refléta un débat plus large, qui confronte encore différentes visions du rôle de l'université (Pachler, Daly, Mor et Mellar, 2010). *In*

fine, des questions furent publiées sur la plateforme d'apprentissage, sous la forme de questionnaires interactifs générés par le logiciel Hot Potatoes¹ (Vincke et Depiereux, 2010). Chaque proposition de réponse cochée par l'étudiant lui renvoyait dans une fenêtre secondaire la raison pour laquelle l'option choisie était correcte ou non. Parallèlement, nous avons développé un logiciel d'enregistrement via le Web des réponses à de petites collections de questions, ce qui nous a permis de mettre en œuvre une stratégie d'évaluation formative encadrée dans le cadre du dispositif d'apprentissage présentiel (Vincke et Depiereux, 2008).

Pour ne pas appauvrir la banque de questions certificatives, celles-ci ont été diversifiées et multipliées par « clonage et mutation ». Le « clonage » consiste à modifier la contextualisation de façon à ce que l'étudiant ne puisse pas se contenter de retenir que si le contexte parle de « moutons », il faut répondre « la corrélation est significative ». La « mutation » consiste à modifier un paramètre de l'énoncé (par exemple $R \leq 0$ est muté en $R \geq 0$), ce qui transforme l'option correcte en distracteur et réciproquement. Cette technique de diversification des questions est particulièrement bien adaptée à notre contexte et permet, dans une stratégie d'évaluation formative, de mobiliser certaines notions de façon répétitive dans différents contextes et de conserver des questions inédites pour l'évaluation certificative. Au cours du temps, la rédaction de questions originales, à leur tour clonées et mutées, a ainsi permis d'enrichir la collection pour atteindre environ 700 questions. Celles-ci sont classées en trois catégories selon qu'elles évaluent a) les connaissances théoriques, b) la maîtrise de techniques ou c) la sélection et l'utilisation efficace de la technique adaptée au regard de l'énoncé de la question et sur la base des connaissances théoriques de l'apprenant (la justification de choix est ce que Chevallard, 1999, appelle un discours technologique).

L'évaluation certificative devant être simultanée pour environ 1 000 étudiants, elle ne peut être réalisée directement sur ordinateur. Une banque de questions indépendante est donc maintenue dans le logiciel FileMaker Pro pour pouvoir sélection-

ner les questions non redondantes entre les sessions d'examens et générer les questionnaires papier.

La procédure que nous avons détaillée jusqu'ici présente plusieurs désavantages : i) la triplification des collections de questions (formative non présente sur Hot Potatoes, formative présente sur le Web et certificative dans FileMaker) implique inévitablement une évolution divergente de celles-ci; ii) aucun des deux logiciels ne permet de conserver l'historique des corrections apportées aux énoncés et options de réponse; iii) Hot Potatoes ne permet pas d'associer les propositions de réponses à des coefficients de certitude; iv) Hot Potatoes ne permet pas de suivre les réponses des étudiants; v) l'analyse rétrospective du profil de réponse aux questions de l'évaluation certificative ne fournit aucune garantie que la question posée au temps t n'a pas été corrigée lorsqu'elle est à nouveau proposée au temps $t + 1$.

Détection d'obstacles cognitifs

Parallèlement à ces développements, nous avons commencé à tenter de diagnostiquer les obstacles cognitifs de la statistique en réalisant des interviews semi-structurées d'étudiants (Calmant, 2004; Calmant, Ducarme et Schneider, 2011; Vincke et al., 2013). Mais ces interviews, bien que très instructives, sont assez lourdes à réaliser, et donc impossibles à généraliser. C'est ainsi que nous nous sommes intéressés aux plateformes QCM, en postulant que l'enregistrement continu et l'analyse des réponses des étudiants pourraient nous permettre de détecter des obstacles cognitifs de façon systématique en déterminant le moment où une majorité décroche au sein d'une séquence de questions de complexité croissante, pénétrant dans un même concept (voir ci-dessous : modèle didactique et figure 2).

L'utilisation d'un logiciel permettant de systématiser cette analyse permettrait de suivre la cinétique de l'apprentissage, de localiser les obstacles cognitifs et de vérifier l'adéquation des performances observées dans les domaines théorique, technique et technologique. Ce dernier point permettrait de vérifier que l'objectif de l'enseignement, que l'étu-

diant s'approprie une praxéologie (dans le sens de Chevallard, 1999) relative à chaque domaine de compétence, soit bien respecté. Le suivi des performances d'étudiants est également visé dans notre démarche, mais une bonne méthodologie a été présentée à ce sujet par Jaillet (2012).

N'ayant trouvé aucun logiciel correspondant aux fonctionnalités recherchées, le développement d'un nouveau logiciel d'évaluation formative fut entrepris dans le projet eTests.

Modèles

Les modèles du dispositif d'apprentissage centré sur l'apprenant ayant été décrits par ailleurs (Vincke et al., 2013), nous nous focalisons ici sur l'articulation enseignement-évaluation. Chacun des modèles justifie nos choix vis-à-vis de l'objet de l'étude. Le modèle pédagogique pose le cadre théorique de notre dispositif hybride présentiel/numérique et le rôle de l'évaluation formative, concept peu développé dans notre contexte universitaire et sur lequel nous focalisons notre étude. Le modèle d'apprentissage situe notre dispositif dans la mouvance socioconstructiviste, ce qui justifie les interactions créées entre étudiants tant dans l'espace présentiel que numérique. Le modèle didactique focalise le processus d'évaluation formative vers une application encore peu développée dans ce contexte : la détection d'obstacles cognitifs. Ces obstacles sont situés dans la classification de Brousseau afin de les traiter spécifiquement dans notre utilisation de la plateforme eTests.

Modèle pédagogique

Selon Allal et Mottier Lopez (2005), le concept d'évaluation formative a été introduit par Scriven (1967) comme un outil permettant d'ajuster les programmes scolaires. Le concept a ensuite été repris par Bloom, qui fut l'un des premiers à le considérer comme un outil d'apprentissage (Bloom, 1968; Bloom, Hasting et Madaus, 1971), plaçant cette évaluation à la fin d'une séquence d'enseignement

et utilisant les résultats pour organiser une remédiation. Depuis, le concept d'évaluation formative s'est peu à peu étoffé (Allal, 1979; Andrade et Cizek, 2010; Black et Wiliam, 2009), son statut d'objet d'apprentissage est reconnu et ce type d'activité est régulièrement intégré au sein même des séquences d'enseignement.

Le modèle pédagogique sur lequel nous avons construit notre dispositif d'apprentissage (Vincke et al., 2013) guide l'apprenant en lui proposant deux espaces d'activité partiellement redondants, un espace numérique, constitué de la plateforme Web et un espace présentiel, représenté par le cours magistral et les travaux pratiques (figure 1).

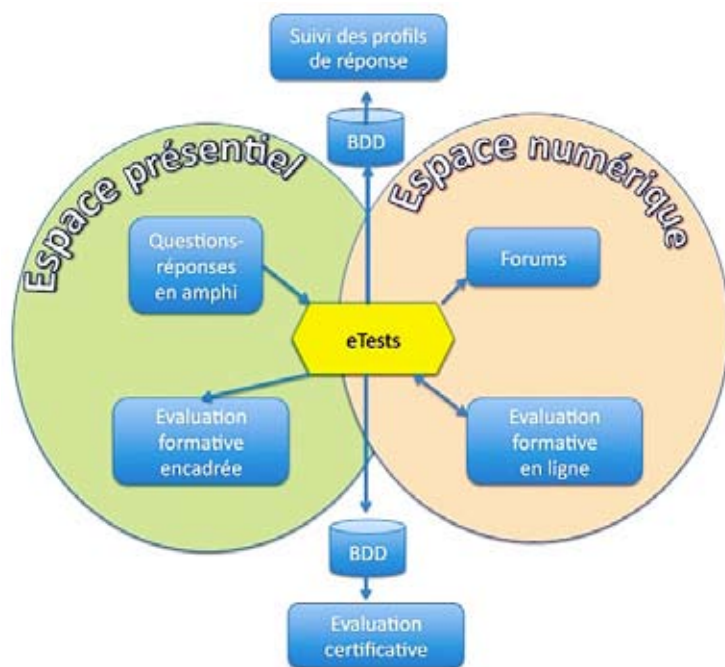


Figure 1. Modèle pédagogique de l'articulation enseignement-évaluation. eTests se situe à l'intersection de deux espaces d'activité, présentiel et numérique, réglant la dévolution des tâches et la gestion du temps didactique. La base de données de questions (BDD) permet de générer les évaluations formatives et certificatives, et la BDD des réponses, le suivi a posteriori des profils de réponse par des indices éducatifs.

L'espace d'activité numérique est un site Web d'autoapprentissage (Vincke et Depiereux, 2010) commun à plusieurs institutions, au sens de Chevallard (1999). Asynchrone, délocalisé et dématérialisé, il correspond en quelque sorte au « cyberprofesseur » de Lombard (2007). L'apprenant y étant beaucoup plus autonome que dans le dispositif présentiel classique (cours ex cathedra, séances de questions-réponses, travaux pratiques encadrés), il favorise la dévolution des tâches au sens de Brousseau (1998) et transforme le temps didactique, comme défini par Chopin (2005).

La coexistence de plusieurs méthodes d'apprentissage procure à chaque étudiant la possibilité d'avancer à son propre rythme, appliquant ainsi les principes de la pédagogie différenciée décrite par Hotte, Basque, Page-Lamarque et Ruelland (2007). Cette approche s'inscrit également dans la pédagogie de la maîtrise (Bloom, 1968) avec des activités présentes suivies d'une évaluation formative dans l'espace d'activité numérique, suivie d'un retour au présentiel pour la remédiation, les deux dernières étapes étant répétées jusqu'à ce que chaque apprenant actif maîtrise la matière abordée.

Modèles d'apprentissage

À l'heure actuelle, la plupart des modèles d'apprentissage font référence aux travaux de deux psychologues : le suisse Jean Piaget, reconnu comme le père du constructivisme, et le russe Lev Vygotski, qui le fait évoluer en socioconstructivisme. Ces deux psychologues font partie d'une même mouvance théorique selon laquelle l'apprenant construit son savoir par son activité propre et au travers des interactions qu'il développe avec autrui. Pour Piaget (1936), l'apprentissage est un phénomène dynamique en constante évolution, durant lequel les acquis sont progressivement reconstruits par l'apprenant. L'apport des travaux de Vygotski (1985) complète ce modèle en y ajoutant

une composante sociale : la reconstruction des acquis en savoir est sensible à l'environnement social dans lequel un apprenant évolue, d'autant plus que ce dernier y établit des interactions significatives. La construction personnelle d'un savoir serait ainsi influencée par l'appropriation éventuelle de ceux d'autrui. Vygotski définit ainsi la zone proximale de développement (ZPD) : la différence entre ce qu'un apprenant peut réaliser seul et ce qu'il peut réaliser avec la collaboration d'un « expert », qui peut aussi bien symboliser une personne isolée ou un groupe d'individus que, selon Lombard (2007), un cyberdispositif.

Dans les modèles de Piaget et de Vygotski, l'apprentissage est présenté comme une activité conflictuelle : conflit entre les conceptions préalables et celles qui émergent d'une analyse logique d'une situation, et conflit entre individus d'avis différents qui argumentent pour défendre leurs positions. Le conflit sociocognitif serait alors le moteur de l'apprentissage (Doise, Mugny et Perret-Clermont, 1975; Perret-Clermont, 1979), permettant à l'apprenant de reconstruire un savoir plus robuste. Suivant cette théorie, chaque membre d'un dispositif d'apprentissage est *de facto* un acteur potentiel de l'apprentissage des autres membres. C'est le fondement des méthodes d'enseignement basées sur le travail collaboratif, dont Bouzidi et Jaillet (2007) poussent la logique jusqu'à l'évaluation par les pairs.

Le dispositif d'apprentissage dans lequel nous utilisons eTests s'inspire de cette mouvance socio-constructiviste. À l'instar de Lombard (2007), nous pensons que le cyberdispositif, avec les interactions qu'il offre sur le forum de la plateforme d'apprentissage (Vincke et Depiereux, 2010) et sur eTests (l'individu situe sa réponse par rapport à celles du groupe, reçoit des justificatifs sur le caractère correct ou incorrect de son choix de réponse et peut discuter de la question asynchroniquement avec des pairs distants), peut être lui-même assimilé à un « expert » et servir de relais vers les autres membres du groupe. Durant les travaux pratiques (espace d'activité présentiel), les apprenants sont installés par binômes devant les ordinateurs, et il leur est conseillé de s'entraider. Chaque partenaire du

binôme dispose de son propre questionnaire eTests sur une moitié d'écran et les réponses suscitent le débat. L'objectif est d'étendre la ZPD à travers leur dialogue et la confrontation de leurs points de vue. Si le succès didactique d'un dispositif d'apprentissage dépend entre autres de la motivation de l'apprenant, il assume lui-même une grande part de responsabilité dans le développement et le maintien de cette motivation (Lebrun, 2007) : dans les faits, l'espace numérique développé par l'équipe enseignante est complété, voire dépassé, par les réseaux sociaux dans lesquels nous n'avons pas (encore?) pris de rôle actif.

Modèle didactique

L'évaluation formative est envisagée comme une série de questions en cascade, pénétrant dans le concept par des questions simples se complexifiant progressivement, ou posant un problème global qui sera par la suite décortiqué par différentes questions plus ciblées abordant les différents niveaux théoriques ou techniques nécessaires pour construire le discours technologique (voir discussion, figure 5). Cette cascade est schématisée à la figure 2.

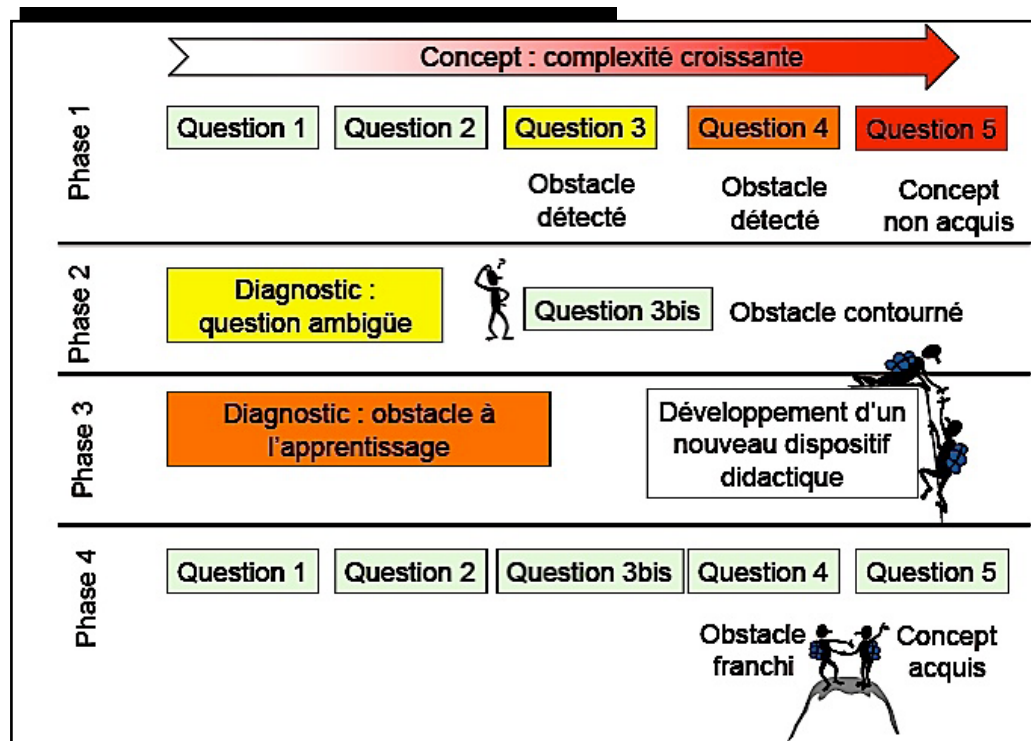


Figure 2. Modèle didactique. Suivi de questions de complexité croissante pénétrant « en cascade » un concept donné (exemple fictif). Vert : plus de 60 % de succès, jaune : entre 60 % et 40 %, orange : entre 40 % et 20 %, rouge : moins de 20 %. Phase 1 : série de questions originales et décrochage des étudiants à partir de la question 3; analyse de la question mettant en évidence son ambiguïté. Phase 2 : modification de la formulation de la question 3 (3bis). Phase 3 : analyse de la question 4 mettant en évidence un obstacle cognitif; une nouvelle séquence d'apprentissage est développée. Phase 4 : performances après implémentation d'une nouvelle séquence d'apprentissage.

L'étudiant rencontre ces questions dans un ordre aléatoire et non dans la séquence de complexité croissante préétablie. L'analyse rétrospective des réponses fournies par l'étudiant entre l'initiation de l'apprentissage et l'évaluation certificative permet par une série de mesures (voir « édumétrie ») : i) de repérer les questions ambiguës ou erronées; le cours se répétant chaque quadrimestre, lorsqu'une question doit être modifiée, il est possible de mesurer à court terme l'impact de ces modifications; ii) de

modifier la séquence d'apprentissage (Vincke et al., 2013) si la distribution des réponses aux questions, avérées correctes et non ambiguës puis reclassées dans la séquence de complexité croissante, témoigne d'une difficulté réelle pour les apprenants, révélant peut-être la présence d'un obstacle cognitif. À plus long terme, l'analyse des fréquences de réponse permettra de vérifier si cet obstacle aura été levé ou non. La mise en œuvre de ce modèle nécessite que la base de données permette la traçabilité des réponses au cours des modifications successives des questions et qu'elle permette de reconstituer le profil des réponses à chaque question en l'état où celle-ci était avant d'éventuelles modifications. Cette fonctionnalité est une des caractéristiques originales de notre logiciel eTests.

La notion d'obstacle et son rôle dans l'apprentissage

La prise en compte d'obstacles à l'apprentissage ainsi que l'aménagement d'une rencontre de l'apprenant avec ceux-ci semblent être fondamentaux. La leçon que l'apprenant tirera de la rencontre avec

l'obstacle sera toujours plus porteuse de sens que celle que le professeur aurait pu lui dispenser avant la réalisation de l'exercice. Le processus d'apprentissage n'est pas un phénomène linéaire, il peut être schématisé comme une progression qui est parfois ralentie, voire stoppée par des obstacles de nature et d'origine variées et qui nécessitent de la part de l'apprenant un effort supplémentaire afin de les franchir. Sans entrer dans les détails, Brousseau (1998) classe ces obstacles cognitifs en plusieurs catégories, selon leur origine : i) les obstacles ontogéniques qui résultent du développement psychogénétique de l'apprenant; ii) les obstacles d'origine didactique dont la responsabilité incombe aux choix didactiques de l'enseignant (éventuellement inévitables, voire intentionnels); iii) les obstacles épistémologiques, au sens de Bachelard (1938), qui trouvent leur origine dans les conceptions préalables de l'apprenant, l'inconscient collectif et la culture ou la connaissance scientifique de toute la société concernée par l'enseignement analysé. La statistique, en particulier, ne manque pas d'obstacles épistémologiques, par exemple l'affrontement conceptuel entre la norme et l'écart à la norme (Quetelet, 1871) fortement marqué par le contexte socioculturel d'une culture donnée (Chevallard, 1978).

Face à un obstacle cognitif, il y a plusieurs postures possibles de la part de l'enseignant, qui peut soit l'ignorer, soit l'affronter par un dispositif adéquat (Brousseau, 1998). Idéalement, le dispositif sera construit de telle sorte qu'à l'issue de cette rencontre, le ou les obstacles en question soient naturellement définis et franchis. Le but majeur de notre utilisation de la plateforme eTests est de réaliser une détection systématique des obstacles, en nous focalisant tout particulièrement sur ceux qui ont une portée didactique et épistémologique, c'est-à-dire ceux contre lesquels nous avons pu (Vincke et al., 2013) et pourrions encore éventuellement agir.

Le logiciel d'évaluation eTests

Implémentation

L'utilisation de systèmes informatisés pour gérer des évaluations formatives n'est pas récente (De Campos, 1990; Leclercq, 1980; Ring, 1993; Séguin, 1984), de même que le choix d'utiliser à cette fin des questions à choix multiples (Callear et King, 1997; Leclercq, 1986, 1987). L'informatisation a permis le traitement de plus grandes quantités d'information, et l'arrivée du Web au début des années 1990 a facilité le développement de stratégies d'évaluations formatives asynchrones (Buchanan, 2000; Gibson, Brewer, Dholakia, Vouk et Bitzer, 1995; Stockburger, 1999; Peat et Franklin, 2002). Mais il est certains domaines qui seront plus difficiles à atteindre avec ce type de média, par exemple lorsqu'il faut prendre en compte d'autres sources d'informations que les réponses à des tests ou encore fournir à chaque étudiant des conseils personnalisés pour orienter son apprentissage, domaines qui, nous l'espérons, feront longtemps encore appel aux relations humaines. Pour pallier cette dualité, nous avons fait le choix d'un dispositif d'apprentissage hybride, alliant les avantages des outils informatiques (rapidité, puissance de calcul), ceux du Web (asynchronicité, personnalisation du rythme d'apprentissage) et ceux d'un encadrement humain (discussion des résultats et remédiation collective ou personnalisée).

Édumétrie

L'édumétrie utilise dans le cadre de l'étude les techniques mathématiques et statistiques d'analyse des réponses développées en psychométrie (Carver, 1974) et se focalise sur l'analyse des scores qui renseignent sur la qualité des questions. Les principaux indicateurs sur lesquels notre logiciel se focalise concernent l'évaluation de la difficulté et de la qualité des questions, en nous inspirant principalement des ouvrages de Leclercq (1987), de Laveault et Grégoire (2002) et de Gilles (2002). L'écran général de l'analyse édumétrique est présenté à la figure 3 et les principaux indicateurs sont détaillés ci-dessous.

Navigation
10 question(s) répartie(s) en 10 pages (1 par page): Page 7/10

Paramètres d'analyse globale de l'enquête

Enregistrer

Considérer : Fréquences r_{pbis}

Réponses : Correctes Incorrectes

Seuils des fréquences des réponses

Inférieur Supérieur

Réponse correcte

Réponse incorrecte

Question 7: Réponse correcte :

Question : Quelle est l'affirmation fautive parmi les suivantes ?

Propositions Réponses : 83 Certitudes r_{pbis}

Question non répondue

1 : La moyenne arithmétique est sensible aux valeurs extrêmes. 1%

2 : L'amplitude est stable dans des petits échantillons issus d'une même population. 51%

3 : Le mode est le centre de la classe la plus fréquente. 25%

4 : Il peut y avoir plusieurs modes dans une distribution. 12%

5 : La médiane est la valeur en dessous de laquelle se trouvent 50% des observations.

Solutions générales :

6 : Toutes les propositions sont correctes 10%

7 : Aucune proposition n'est correcte 1%

8 : Il y a une absurdité dans l'énoncé

9 : Il manque des données pour pouvoir répondre

Identification de la question

Code Version Du Au Utilisateurs

1.1 23/05/2010 En 83

20 cours

Statistiques propres à cette question

Réponses	Total
Correctes	42
Incorrectes	41
OMises	5
OC	88
ONC	83

Indice OC ONC

ρ 0.48 0.51

ρ' 0.41 0.44

$\sqrt{\rho q}$ 0.50 0.49

D_{pool} 0.70 0.66

D_{cert} 0.57 0.58

$D_{no\ cert}$ 0.65 0.67

Fidélité 1.03

r_{pbis} r_{pbis}^2 $r_{pbis}S$ $r_{pbis}SC$ $NCSq$ NCS_t

0.52 -0.12 -0.02 0.01 -0.03 0.36

Aide collaborative

Commentaires (0)

Solutions proposées (0)

Opinion des étudiants sur cette question

Figure 3. Tableau de bord général de l'analyse éduométrique d'une question par le logiciel eTests, tel qu'accessible aux enseignants. Les items relatifs au questionnaire (A) et à la question (B) sont expliqués dans le texte.

Des pictogrammes (A1) récapitulent l'ensemble du questionnaire en pointant les questions qui requièrent une attention particulière. Un graphique (A2), dont l'interprétation est détaillée à la figure 4, présente une synthèse intéressante de la relation entre, en abscisse, p , la fréquence de sélection de la bonne réponse, et en ordonnée, D , un indice de discrimination (Ebel, 1965; Feldt, 1961; Findley, 1956; Kelley, 1939).

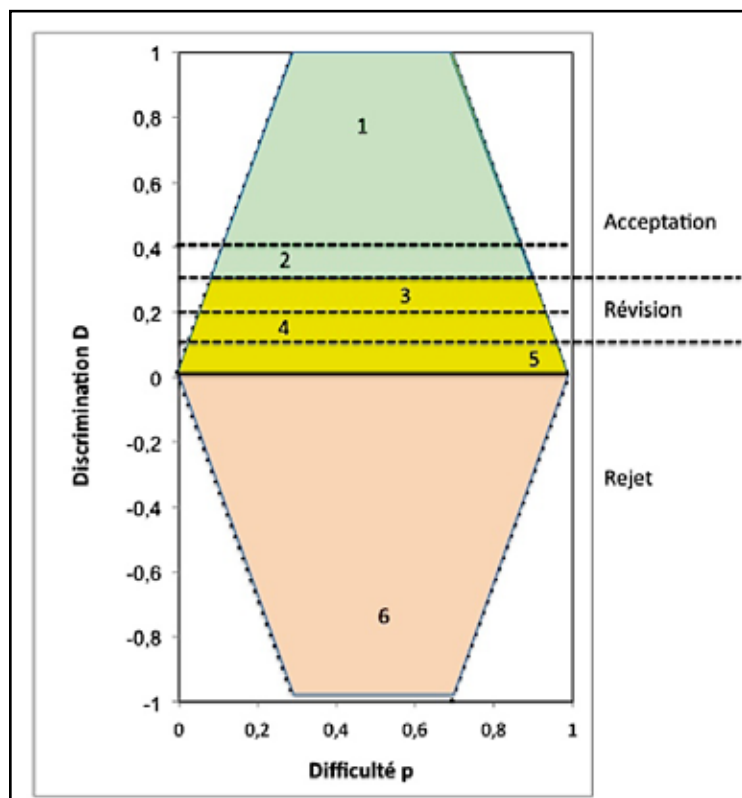


Figure 4. Représentation graphique (Ebel, 1965) de la discrimination (en ordonnée) et de la difficulté (en abscisse) de la question, l'indice p étant faible lorsque la difficulté est élevée et réciproquement. Les exemples et leur explication sont inspirés de ceux de Laveault et Grégoire (2002). 1) Question idéale pour différencier les répondants : difficulté moyenne et discrimination élevée. 2) Question plus difficile et encore discriminante. Questions à examiner : 3) Question de difficulté moyenne qui discrimine assez peu. 4) Question plus difficile qui discrimine assez peu. 5) Question facile et peu discriminante. 6) Question réellement problématique : car la question est de difficulté moyenne, mais discrimine de manière très négative. L'analyse pointe soit une erreur dans la grille de correction ou dans la question, soit un problème d'apprentissage.

L'indice p est faible lorsque la difficulté est élevée et réciproquement (p proche de 0 = difficulté élevée, p proche de 1 = question facile). Idéalement, il aurait dû s'appeler indice de facilité, mais l'expression indice de difficulté semble consacrée dans la littérature (Leclercq et Gilles, 2003). L'indice de discrimination D est facile à calculer et permet de diagnostiquer des problèmes internes aux questions. Selon Leclercq (1987, cité par Gilles, 2002) l'indice D tend vers le coefficient de corrélation point bisériale, ou r_{pbis} (figure 3, A3) (Lord, Novick et Birnbaum, 1968), lorsque le nombre d'observations augmente. Un indice D et/ou un r_{pbis} négatif (zone rouge) signifient que ceux qui ont bien répondu à cette question sont majoritairement ceux qui ont le moins bien répondu à l'ensemble des autres questions du questionnaire, ce qui n'est pas normal (figure 4) (Laveault et Grengoire, 2002) et généralement le signe d'un problème de formulation de la question posée. Une question située dans la zone jaune devrait être examinée : son pouvoir discriminant reste positif, mais faible. Une question située dans la zone verte est une question qui ne présente pas de problème.

Les indicateurs B détaillent les propositions de réponse d'une question particulière. En B1 se trouve une analyse des réponses enregistrées pour chaque proposition, du haut vers le bas : réponse correcte (identifiée par la flèche verte), distracteurs et solutions générales. B1a est la fréquence des réponses, B1b est la proportion des coefficients de certitude (jaune : peu certain; vert clair : assez certain; vert foncé : tout à fait certain), B1c est le r_{pbis} calculé par proposition de réponse. Ces trois indicateurs sont récapitulés dans un pictogramme semblable à A1 mais ici spécifique de chaque proposition de réponse, ce qui met en évidence au moins un problème dans l'option par un examen visuel rapide.

Dans la question illustrée à la figure 3, la proposition correcte (n° 2) n'est choisie que par 51 % des 83 répondants avec une confiance mitigée. Par contre, la proposition n° 3 est détectée comme un distracteur attractif, avec 25 % des réponses, dont la moitié avec une certitude élevée.

En B2 et B3 sont repris différents indices complémentaires dont la comparaison des valeurs relève plus d'une recherche méthodologique que de la simple analyse du questionnaire et dont la discussion de la spécificité dépasse le cadre de cet article.

Caractéristiques techniques

Les aspects techniques du logiciel n'ont pas grand intérêt dans cette discussion. Nous en mentionnons ici les points essentiels. eTests est une application Web, diffusée sous licence EUPL V1.1. Le code source est téléchargeable sur sourceforge.net² d'où il a déjà été téléchargé plus de 600 fois depuis plus de 90 pays différents. Le code est écrit en PHP³, et il dialogue avec une base de données MySQL⁴. La version actuelle (0.3.1) est stable (elle est utilisée dans deux universités belges depuis trois ans par 1 000 à 3 000 utilisateurs/an) et son interface est actuellement accessible en français et en anglais. À la suite de la présentation d'eTests au FOSDEM 2013⁵, des contacts ont été pris avec des développeurs brésiliens et indiens pour étendre la communauté et accélérer son développement.

Discussion

L'objectif de l'évaluation formative est de responsabiliser l'étudiant vis-à-vis de son apprentissage en lui permettant de tester ses performances, de les comparer à celles de ses pairs et d'en observer l'évolution, en fonction du temps et de l'effort d'étude consenti dans un laps de temps donné. Contrairement aux plateformes étudiantes classiques – Moodle⁶, Claroline (Lebrun, 2004)... – dont la structure est une série de cours indépendants les uns des autres au sein d'un établissement d'enseignement unique, eTests propose des espaces de partage de contenu évaluatif entre plusieurs titulaires pouvant appartenir à des établissements différents. Ce logiciel permet donc une meilleure intégration des enseignements et la répartition de l'effort que représente la réalisation d'une base de questions couvrant l'ensemble d'une matière.

Toute modification de question entraîne automatiquement l'enregistrement complet de la question modifiée sous un numéro de version incrémenté, sans suppression de ses caractéristiques précédentes. Cette fonctionnalité permet d'analyser les résultats issus de questionnaires en fonction de l'état réel d'une question au moment où elle a été utilisée. Il n'existe pas, à notre connaissance, d'autre logiciel libre permettant d'obtenir ce suivi des questions ou une analyse d'une grande précision en fonction d'un relevé éduométrique très diversifié. Cela rend le logiciel particulièrement utile pour assurer la localisation et le suivi des obstacles cognitifs.

L'analyse des données enregistrées par ce système a permis la mise en évidence de phénomènes inattendus. Les différentes notions qui présentent manifestement des difficultés n'étaient pas forcément celles que les enseignants attendaient. La figure 5 illustre la détection d'un obstacle cognitif par une cascade de quatre questions dans l'évaluation de la tâche « calculer une variance ».

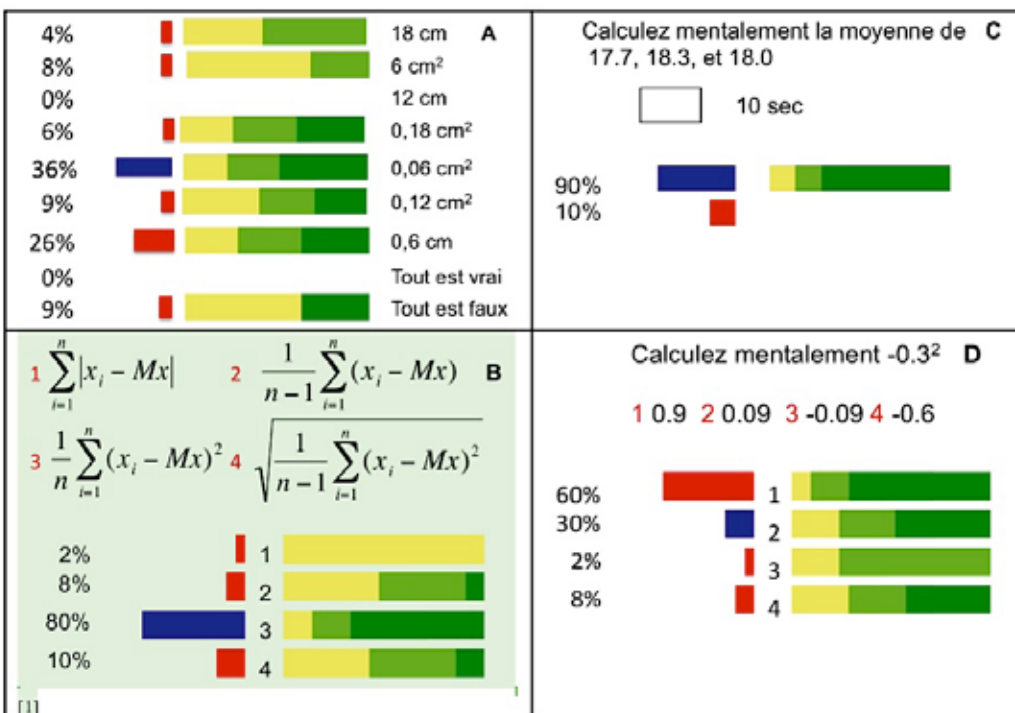


Figure 5. Profils de réponse à une série de questions en cascade. À gauche : p, proportion de bonnes réponses. À droite : proportion des coefficients de certitude. A. Type de tâche : « calculer une variance » (voir texte); B. Choix de 4 formules de calcul alternatives (approche théorique); C. et D. Étapes de calcul mental (approche technique). L'interprétation est fournie dans le texte.

La première question consiste à calculer mentalement la variance d'un échantillon. « Lors d'une récolte à la pompe à vers sur une plage de Bretagne, un étudiant récolte trois *Lanice conchilega*. Ces vers mesurent respectivement 17,7, 18,3 et 18,0 cm. Quelle est la valeur de la variance de cet échantillon de vers? » Le taux de bonnes réponses est faible et la certitude des étudiants, mitigée. Sur la base de cette seule question, le diagnostic hâtif se limiterait au fait que le concept statistique de variance n'est pas maîtrisé.

La seconde question propose un choix de quatre formules de calcul ressemblantes, dont seule la troisième est correcte : le taux de bonnes réponses et la certitude des étudiants sont élevés; la couche théorique est donc maîtrisée.

La troisième question demande le calcul mental rapide de la moyenne de l'échantillon : le taux de bonnes réponses et la certitude des étudiants sont élevés; le problème ne vient donc pas de cette compétence technique.

La quatrième question demande le calcul mental d'une étape du calcul de la variance :

$-0,3^2$; le taux de bonnes réponses est faible et la certitude associée à une fausse réponse est élevée. L'analyse montre donc que l'étape limitante du problème complexe est la capacité de l'étudiant d'élever au carré une valeur négative et inférieure à 1, prérequis qui relève du cours de mathématique. L'obstacle est donc détecté, et peut faire l'objet d'une remédiation particulière.

La mise en place de notre dispositif a été très appréciée des étudiants. Lors de l'évaluation institutionnelle des enseignements, gérée anonymement sur la plateforme étudiante de notre établissement⁷, et lors d'enquêtes indépendantes (Vincke, 2010), ils ont plébiscité à plusieurs reprises ce système, qui selon eux est déterminant tant pour leur apprentissage que pour la réussite de l'examen. Lors d'un travail en tandem dans un dispositif encadré, chaque étudiant doit entrer sa propre réponse et son coefficient de certitude. Ainsi, plutôt que de se contenter de copier l'un sur l'autre, les étudiants ont

rapidement pris conscience de leur intérêt à répondre honnêtement au questionnaire. Cette forme de compétition, sans autre enjeu que celui de l'apprentissage, est une réelle émulation. Nous assistons à une meilleure appropriation par les étudiants du système d'évaluations, qui s'est d'ailleurs traduite par une très nette baisse des réactions manifestant un stress vis-à-vis de la méthodologie de l'examen certificatif de fin d'année.

Conclusions et perspectives

Vu l'intérêt des étudiants et la richesse des échanges générés avec les enseignants, une mise en ligne de QCM, qui n'avait au départ qu'un but de démythification de l'évaluation certificative, a été progressivement généralisée et a évolué vers un support de l'apprentissage par des évaluations formatives. Pour systématiser cette approche et développer une base de données cohérente de questions formatives, nous avons développé le logiciel eTests.

L'analyse rétrospective de la base de données donne à l'enseignant la possibilité du suivi des questions. Des profils anormaux permettent en outre de détecter des difficultés qui peuvent se révéler être des obstacles didactiques ou épistémologiques auxquels il peut être possible de remédier.

La détection et la tentative de résolution des obstacles didactiques se déroulent donc selon le schéma suivant : i) une séquence de questions de complexité croissante visant un concept déterminé est choisie dans la base de données ou élaborée si elle est inexistante; ii) l'étudiant rencontre ces questions dans un ordre aléatoire; iii) le suivi des performances des apprenants permet de déterminer les questions qui correspondent à des obstacles qu'ils n'ont pas surmontés; iv) l'analyse des indices éducatifs permet de déterminer si le problème vient de la formulation de la question (ambiguë, mauvais choix des distracteurs, bonne réponse discutable, question prématurée vis-à-vis de prérequis à ce stade de l'apprentissage...) ou de l'appropriation du concept; v) dans ce dernier cas, une séquence d'apprentissage est imaginée pour permettre de

contourner ou de surmonter l'obstacle en question (Vincke et al., 2013).

Nous développons actuellement une séquence de questions relative à un obstacle épistémologique qui concerne la compréhension des risques d'erreurs dans les tests d'hypothèse. Cet obstacle est lié à la confusion historique des notions d'erreur et de variabilité, comme dans des expressions consacrées telles qu'« erreur type » (Depiereux, 2013). Nous suspectons par ailleurs que la compréhension de ce concept soit liée aux croyances épistémiques des étudiants (Schommer-Aikins et Duell, 2013). En bref, le modèle théorique oppose une perspective naïve de la connaissance (immuable et détenue par l'autorité) à un point de vue sophistiqué (évolution de la connaissance et de l'aptitude à apprendre, interconnexion des concepts) (Huet et Escribe, 2004). Notre objectif est d'étudier la corrélation éventuelle entre le profil de réponse, récolté sur eTests, et le profil de croyances épistémiques récolté dans une version française du questionnaire des croyances épistémiques de Schommer.

Références

- Albero, B. (2011). Le couplage entre pédagogie et technologies à l'université : cultures d'action et paradigmes de recherche. *Revue internationale des technologies en pédagogie universitaire*, 8(1-2), 11-21. [Récupéré de http://www.ritpu.org](http://www.ritpu.org)
- Allal, L. (1979). Stratégies d'évaluation formative : conceptions psychopédagogiques et modalités d'application. Dans L. Allal, J. Cardinet et P. Perrenoud (dir.), *L'évaluation formative dans un enseignement différencié* (p. 153-183). Berne, Suisse : Peter Lang.
- Allal, L. et Mottier Lopez, L. (2005). L'évaluation formative de l'apprentissage : revue de publications en français. Dans OCDE (dir.), *L'évaluation formative. Pour un meilleur apprentissage dans les classes* (p. 265-290). Paris, France : auteur. [10.1787/9789264007420-5-fr](http://dx.doi.org/10.1787/9789264007420-5-fr)
- Andrade, H. J. et Cizek, G. J. (dir.). (2010). *Handbook of formative assessment*. New York, NY : Routledge.
- Bachelard, G. (1938). *La formation de l'esprit scientifique*. Paris, France : Librairie Philosophique J. Vrin.
- Black, P. et Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21(1), 5-31. doi:10.1007/s11092-008-9068-5
- Bloom, B. S. (1968). Learning for mastery. *Evaluation Comment*, 1(2), 1-12.
- Bloom, B. S., Hasting, J. T. et Madaus, G. F. (1971). *Handbook on formative and summative evaluation of student learning*. New York, NY : McGraw-Hill.
- Bouzidi, L. et Jaillet, A. (2007). L'évaluation par les pairs pourra-t-elle faire de l'examen une vraie activité pédagogique? Dans *Actes de la conférence Environnements informatiques pour l'apprentissage humain (EIAH)*. [Récupéré del'archive HAL : http://hal.archives-ouvertes.fr](http://hal.archives-ouvertes.fr)
- Brousseau, G. (1998). *Théorie des situations didactiques*. Grenoble, France : La pensée sauvage.
- Buchanan, T. (2000). The efficacy of a worldwide web mediated formative assessment. *Journal of Computer Assisted Learning*, 16(3), 193-200. doi:10.1046/j.1365-2729.2000.00132.x
- Callear, D. et King, T. (1997). Using computer-based tests for information science. *Research in Learning Technology*, 5, 27-31. [Récupéré de http://www.researchinlearningtechnology.net](http://www.researchinlearningtechnology.net)
- Calmant, P. (2004). *Favoriser l'apprentissage des biostatistiques par le Web? Essai de problématisation d'une question issue du terrain* (thèse de doctorat non publiée). Facultés universitaires Notre-Dame de la Paix, Namur, Belgique.

- Calmant, P., Ducarme, M. et Schneider, M. (2011). Obstacles a priori à l'apprentissage de l'analyse statistique inférentielle. *Statistique et enseignement*, 2(1), 43-59. [Récupéré de http://publications-sfds.math.cnrs.fr/index.php/StatEns](http://publications-sfds.math.cnrs.fr/index.php/StatEns)
- Carver, R. P. (1974). Two dimensions of tests: Psychometric and edumetric. *American Psychologist*, 29(7), 512-518. doi:10.1037/h0036782
- Chevallard, Y. (1978). *Problèmes de surdétermination en didactique : la notion de moyenne en statistique*. Document inédit, Laboratoire de psychologie du travail du Cnam, département de mathématique, UER de Luminy et IREM d'Aix-Marseille.
- Chevallard, Y. (1999). L'analyse des pratiques enseignantes en théorie anthropologique du didactique. *Recherches en didactique des mathématiques*, 19(2), 221-265.
- Chopin, M.-P. (2005, octobre). *Le temps didactique en théorie anthropologique du didactique : quelques remarques méthodologiques à propos des moments de l'étude*. Communication présentée au 1^{er} congrès international sur la Théorie anthropologique du didactique [TAD], « Société, école et mathématiques : apports de la TAD », Baeza, Espagne. [Récupéré du site du congrès : http://www4.ujaen.es/~aestepa/TAD-frances](http://www4.ujaen.es/~aestepa/TAD-frances)
- De Campos, M. (1990). Outils diagnostiques et enseignement assisté par ordinateur. *Mesure et évaluation en éducation*, 13(3), 55-69.
- Depiereux, E. (2013). *De la variabilité aux risques d'erreur. Analyse critique des résultats expérimentaux et de la fiabilité d'une décision clinique*. Namur, Belgique : Presses universitaires de Namur.
- Doise, W., Mugny, G. et Perret-Clermont, A. N. (1975). Social interaction and the development of cognitive operations. *European Journal of Social Psychology*, 5(3), 367-383. doi:10.1002/ejsp.2420050309
- Ebel, R. L. (1965). *Measuring educational achievement*. Englewood Cliff, NJ : Prentice Hall.
- Feldt, L. S. (1961). The use of extreme groups to test for the presence of a relationship. *Psychometrika*, 26(3), 307-316. doi:10.1007/BF02289799
- Findley, W. G. (1956). A rationale for evaluation of item discrimination statistics. *Educational and Psychological Measurement*, 16(2), 175-180. doi:10.1177/001316445601600201
- Gibson, E. J., Brewer, P. W., Dholakia, A., Vouk, M. A. et Bitzer, D. L. (1995). *A comparative analysis of web-based testing and evaluation systems* (rapport n° 27695-8206). Raleigh, NC: North Carolina State University. [Récupéré du site CiteSeerX : http://citeseerx.ist.psu.edu](http://citeseerx.ist.psu.edu)
- Gilles, J.-L. (2002). *Qualité spectrale des tests standardisés universitaires* (thèse de doctorat non publiée). Université de Liège, Belgique.
- Hotte, R., Basque, J., Page-Lamarche, V. et Ruelland, D. (2007). Ingénierie des compétences et scénarisation pédagogique. *Revue internationale des technologies en pédagogie universitaire*, 4(2) 38-56. [Récupéré de http://ritpu.org](http://ritpu.org)
- Huet, N. et Escribe, C. (2004). Croyances épistémiques, buts d'accomplissement de soi et engagement dans l'utilisation d'un média électronique chez des étudiants. *Revue des sciences de l'éducation*, 30(1), 177-196. doi:10.7202/011776ar
- Jaillet, E. (2012). C2i2e à distance : l'activité de l'étudiant est-elle représentative de son évaluation? *Revue internationale des technologies en pédagogie universitaire*, 9(1-2), 100-110. [Récupéré de http://ritpu.org](http://ritpu.org)
- Kelley, T. L. (1939). Selection of upper and lower groups for the validation of test items. *Journal of Educational Psychology*, 30(1), 17-24. doi:10.1037/h0057123
- Laveault, D. et Grégoire, J. (2002). *Introduction aux théories des tests en psychologie et en sciences de l'éducation* (2^e éd.). Bruxelles, Belgique : De Boeck.

- Lebrun, M. (2004). Enseigner et apprendre en ligne. Claroline et le site iCampus de l'UCL : fondements, outils, dispositifs. *Les dossiers des sciences de l'éducation*, 12, 79-93.
- Lebrun, M. (2007). *Des technologies pour enseigner et apprendre*. Bruxelles, Belgique : De Boeck.
- Leclercq, D. (1980). Computerized tailored testing: Structured and calibrated item banks for summative and formative evaluation. *European Journal of Education*, 15(3), 251-260.
- Leclercq, D. (1982). Confidence marking: Its use in testing. *Evaluation in Education*, 6(2), 161-287.
- Leclercq, D. (1986). *La conception des questions à choix multiple*. Bruxelles, Belgique : Labor.
- Leclercq, D. (1987). *Qualité des questions et signification des scores (avec application aux QCM)*. Bruxelles, Belgique : Labor.
- Leclercq, D. et Gilles, J.-L. (2003). Analyses psychométriques des questions des 10 check-up MOHICAN : vue d'ensemble. Dans D. Leclercq (dir.), *Diagnostic cognitif et métacognitif au seuil de l'université : le projet MOHICAN mené par les 9 universités de la Communauté française Wallonie-Bruxelles* (p. 173-180). Liège, Belgique : Éditions de l'Université de Liège.
- Lombard, F. (2007). Du triangle de Houssaye au tétraèdre des TIC : comprendre les interactions entre les savoirs d'expérience et ceux de recherche. Dans B. Charlier et D. Peraya (dir.), *Transformation des regards sur la recherche en technologie de l'éducation* (p. 137-155). Bruxelles, Belgique : De Boeck.
- Lord, F. M., Novick, M. R. et Birnbaum, A. (1968). *Statistical theories of mental test scores*. Oxford, R.-U. : Addison-Wesley.
- Pachler, N., Daly, C., Mor, Y. et Mellar, H. (2010). Formative e-assessment: Practitioner cases. *Computers & Education*, 54(3), 715-721. doi:10.1016/j.compedu.2009.09.032
- Peat, M. et Franklin, S. (2002). Supporting student learning: The use of computer-based formative assessment modules. *British Journal of Educational Technology*, 33(5), 515-523. doi:10.1111/1467-8535.00288
- Perret-Clermont, A. N. (1979). *La construction de l'intelligence dans l'interaction sociale*. Berne, Suisse : Peter Lang.
- Piaget, J. (1936). *La naissance de l'intelligence chez l'enfant*. Neuchâtel, Suisse : Delachaux et Niestlé.
- Quetelet, A. (1871). *Anthropométrie ou mesure des différentes facultés de l'homme*. Bruxelles, Belgique : C. Muquardt.
- Ring, G. (1993). Characteristics of state-of-the-art computer-administered testing systems. *Australian Educational Computing*, 8(2), 25-28.
- Schommer-Aikins, M. et Duell, O. K. (2013). Domain specific and general epistemological beliefs. Their effects on mathematics. *Revista de Investigación Educativa*, 31(2), 317-330. Récupéré de <http://revistas.um.es/rie>
- Scriven, M. (1967). *Perspectives of curriculum evaluation*. Chicago, IL : Rand McNally.
- Séguin, S. P. (1984). L'utilisation des micro-ordinateurs pour l'évaluation des apprentissages : quelques perspectives d'avenir. *Mesure et évaluation en éducation*, 7(1), 53-66.
- Stockburger, D. (1999). Automated grading of homework assignments and tests in introductory and intermediate statistics courses using active server pages. *Behavior Research Methods, Instruments, & Computers*, 31(2), 252-262. doi:10.3758/BF03207717
- Van Vyve-Genette, A., Gohy, J. M. et Feytmans, E. (1988). *Statistique élémentaire pour les sciences bio-médicales : apprentissage par la micro-informatique*. Bruxelles, Belgique : De Boeck Université.

Vincke, G. (2010). *Évaluation formative en ligne pour favoriser l'apprentissage des biostatistiques. Analyse d'une pratique et identification des facteurs d'apprentissage développés dans la plate-forme eTests* (Mémoire de master, Université Catholique de Louvain). [Récupéré](http://dial.academielouvain.be) du dépôt Institutionnel de l'Académie Louvain : <http://dial.academielouvain.be>

Vincke, G. et Depiereux, E. (2008). *Exemple d'utilisation d'auto-évaluation formative en support de l'apprentissage des biostatistiques en premier cycle universitaire, ou quand l'évaluation des étudiants conduit à la remise en question du système d'apprentissage lui-même*. Communication présentée au 20^e colloque de l'Association pour le développement des méthodologies d'évaluation en éducation, Genève, Suisse. [Récupéré](https://plone2.unige.ch/admee08) du site du colloque : <https://plone2.unige.ch/admee08>

Vincke, G. et Depiereux, E. (2010). Mutation d'un cours de biostatistique : auto-apprentissage sur le Web, autoévaluation, transformation du contrat didactique et diminution du taux d'échec. *Revue internationale des technologies en pédagogie universitaire*, 7(3), 6-12. [Récupéré](http://www.ritpu.ca) de <http://www.ritpu.ca>

Vincke G., Wauthy, A.-C., Bihin, B. et Depiereux, E. (2013). Quand la délocalisation numérique d'une partie d'un dispositif d'apprentissage permet de recentrer le temps présentiel sur un obstacle : exemple de l'appropriation de la courbe de Gauss par la manipulation d'objets concrets. *Revue internationale des technologies en pédagogie universitaire* 10(1), 16-28. [Récupéré](http://www.ritpu.org) de <http://www.ritpu.org>

Vygotski, L. (1985). *Pensée et langage* (F. Sève, trad.). Paris, France : Messidor, Éditions sociales. (Ouvrage original publié en 1933).

Note des auteurs

Sur le plan conceptuel et technique, ce travail qui s'étale sur près de 15 ans est intimement lié à d'autres développements de produits didactiques multimédias. Il est impossible de citer nommément toutes les personnes qui s'y sont impliquées directement ou indirectement. Qu'elles se sentent associées à la courte liste des personnes citées ici : Philippe Calmant, Pierre Dagnelie, Ernest Feytmans, Annick Van Vyve-Genette, Benoit De Hertogh, Isabelle Housen, Marcel Lebrun, Isabelle Motte, Marcel Remon, Daniel Rousselet, Jean-Pierre Troussart ...

Notes

- 1 <http://hotpot.uvic.ca>
- 2 <https://sourceforge.net/projects/etests>
- 3 <http://www.php.net>
- 4 <http://www.mysql.fr>
- 5 Free and Open Source Developers European Meeting, Bruxelles, Belgique, 2-3 février 2013.
- 6 <https://moodle.org>
- 7 À l'affirmation « Le site est une aide précieuse à l'apprentissage », 90 % répondent « totalement d'accord », alors que d'autres affirmations récoltent des scores nettement moins favorables.