

Contribution au symposium n° : [ne pas compléter]

Evaluation et technologies numériques : quels développements futurs pour quelles finalités ?

Prénom et nom de l'auteur 1 : Vinciane Crahay

Attache institutionnelle de l'auteur 1 : SMART-IFRES, Université de Liège

Email pour de l'auteur 1 : v.crahay@ulg.ac.be

Prénom et nom de l'auteur suivant : Pierre Bonnet

Attache institutionnelle de l'auteur suivant : Faculté de Médecine, Université de Liège

Email pour de l'auteur suivant : p.bonnet@ulg.ac.be

Prénom et nom de l'auteur suivant : Aurélie Piazza

Attache institutionnelle de l'auteur suivant : SMART-IFRES, Université de Liège

Email pour de l'auteur suivant : a.piazza@ulg.ac.be

Prénom et nom de l'auteur suivant : Jean-Philippe Humblet

Attache institutionnelle de l'auteur suivant : SMART-IFRES, Université de Liège

Email pour de l'auteur 1 : jp.humblet@ulg.ac.be

Prénom et nom de l'auteur suivant : Valérie Defaweux

Attache institutionnelle de l'auteur suivant : Faculté de Médecine, Université de Liège

Email pour de l'auteur suivant : valerie.defaweux@ulg.ac.be

Prénom et nom de l'auteur suivant : Adrienne Dernier

Attache institutionnelle de l'auteur suivant : Faculté de Médecine, Université de Liège

Email pour de l'auteur suivant : adrienne.dernier@ulg.ac.be

Questions ouvertes : améliorer la fidélité de la correction à l'aide de grilles d'évaluation critériées et l'utilisation d'un logiciel dédié.

Un résumé court (max 300 mots)

Évaluer au travers de questions à réponse ouverte longue, dans le contexte de massification des étudiants devient difficilement réalisable.

Cependant, ce type d'évaluation reste indispensable pour évaluer des performances complexes. Le Système Méthodologique d'Aide à la Réalisation de Tests (SMART) de l'Université de Liège (ULiège) a donc tenté de concilier les besoins purement logistiques d'aide à la correction des enseignants avec les préceptes docimologiques d'une évaluation de qualité.

Pour ce faire, un outil de correction dédié a été développé et est proposé aux enseignants. Il a comme particularités techniques de permettre de dématérialiser les copies, de corriger en ligne (une connexion internet suffit) et d'automatiser le calcul des notes ainsi que leur envoi dans le bulletin de l'étudiant.

Du côté docimologique, l'enseignant réalise la correction à l'aide d'une grille d'évaluation composée d'échelles descriptives, permettant de rendre cette correction la plus objective et la plus reproductible possible.

Cette communication se centre sur l'analyse de la fidélité intra-correcteur des notes attribuées. Notre objectif est d'analyser si l'objectivisation et la reproductibilité de la correction sont rencontrées.

Nous présenterons une expérimentation en trois phases :

- Une première correction ;
- Une seconde correction, deux ans plus tard, des mêmes copies, intégrées parmi les copies de nouveaux étudiants ;
- Une troisième correction, réalisée quatre ans plus tard, suite à laquelle une possibilité de reviewing a été permise (revisualisation de copies en fonction de la note initialement attribuée).

Les trois notes obtenues par les étudiants seront comparées afin d'analyser la fidélité du correcteur et si l'étape de reviewing apporte un avantage supplémentaire.

3 à 5 mots – clés : Question à réponse longue, Logiciel de correction, Fidélité intra-correcteur

Un résumé long (max 1000 mots)

Le Système Méthodologique d'Aide à la Réalisation de Tests (SMART) de l'Université de Liège (ULiège) a pour mission principale d'aider les enseignants à évaluer leurs étudiants. Pour ce faire, nous leur proposons des accompagnements de type méthodologique et logistique s'appuyant sur une démarche de qualité, clairement formalisée et publiée (Gilles, Detroz, Crahay, 2011 ; Detroz, 2016), et soutenue par des outils d'évaluation développés en interne.

Un de nos services est de proposer à nos enseignants un système de correction *on-line* des questions ouvertes. Notre objectif est double :

- faciliter la fastidieuse étape de correction dans un contexte où la massification des étudiants est bien présente ;
- réaliser une évaluation valide, en cohérence avec les objectifs d'apprentissage et le dispositif d'enseignement.

Concrètement, les réponses des étudiants sont rédigées dans un cadre défini sur un formulaire à lecture optique. Celles-ci sont ensuite scannées. Chacune de ces images est reliée à l'identifiant de l'étudiant. Le format de réponse, selon la consigne donnée, peut être un texte à rédiger, un schéma à créer, un schéma à annoter, un graphique à réaliser... De plus, il est demandé à l'enseignant de fournir une grille de correction. Il est vivement conseillé que l'outil d'appréciation utilisé pour évaluer les productions des étudiants soit une grille d'évaluation à échelles descriptives (Scallon, 2004). Une fois les éléments associés, l'enseignant se connecte via internet au logiciel Sqore pour réaliser ses corrections. À l'écran sont présentées la réponse anonymisée d'un étudiant ainsi que la grille de correction. Il ne reste plus qu'à scorer chaque critère avant de passer à la copie suivante. Une fois que toutes les copies sont corrigées, le test peut être validé et les résultats libérés.

Cette nouvelle plateforme aide les enseignants d'un point de vue logistique. La dématérialisation et la traçabilité des copies ainsi que le calcul et le transfert automatiques des notes vers le système de gestion de notes de l'institution soulagent grandement la tâche du correcteur. Cependant, la plus-value qui nous intéresse est de faire entrer ce type de correction dans une démarche docimologique de qualité, notamment en imposant l'utilisation d'une grille

de correction critériée ayant pour vertu d'améliorer l'objectivité des corrections réalisées et d'éviter un certain nombre de biais tel que l'effet de distribution forcée (ajuster le niveau de ses appréciations des performances des étudiants de façon à conserver, d'année en année, approximativement la même distribution (gaussienne) des notes (loi de posthumus)).

L'anonymisation des copies permet quant à elle de palier les biais suivants :

- l'effet du nom : l'anonymisation des copies permet de ne pas être influencé par la consonance du nom-prénom (Babad, 1980 ; Fajardo, 1985 ; Amigues, Bonniol, Caverni, 1975)
- l'effet de stéréotypie : le fait de ne pas savoir qui est l'étudiant que l'on corrige permet de ne plus influencer la note attribuée.

Cette communication se concentre sur l'utilisation de notre plateforme de correction dans le cadre d'un schéma à annoter. Elle a pour objet de vérifier la reproductibilité de la notation d'un correcteur (fidélité intra-juge) (AERA-APA-NCME, 1999) réalisée à l'aide d'une grille d'évaluation critériée. Le but est de s'assurer que si le test est corrigé à plusieurs reprises, les scores attribués aux étudiants restent identiques, démontrant l'objectivité de la grille de correction et de ses critères associés.

Pour ce faire, nous avons procédé comme suit :

Tout d'abord, les copies sont corrigées une première fois dans Sqore. Après un laps de temps significatif, nous proposons à nouveau le lot de copies au même correcteur, mais cette-fois, elles sont dispersées dans un autre tas de copies administrées ultérieurement et dont la question est identique.

S'ensuit une première analyse des résultats (corrélation entre la première et la deuxième correction) qui mènera à l'analyse plus en profondeur des copies qui obtiennent des scores significativement différents (les copies qui passent d'une réussite à un échec et inversement, mais aussi les copies ayant obtenu un score beaucoup plus ou moins élevé entre les deux corrections).

Suite à ce premier constat, nous vérifierons ensuite une série d'autres éléments :

- La grille de correction est-elle assez diagnostique ?
- L'ordre de correction des copies a-t-il eu un impact ? Où est située la copie dont le score est différent ? Se trouve-t-elle après une/des copies jugée(s) médiocre(s), une/des copies jugée(s) excellente(s) ? (Effet de contraste, Bonniol et Piolat, 1971)
- Ces copies ont-elles une bonne qualité formelle (écriture, soin), orthographe, grammaire ?

Après avoir répondu à ces quelques questions, nous soumettons une troisième fois les copies mais en y intégrant une étape de reviewing. Celle-ci consiste, une fois les corrections terminées, en la possibilité de visualiser certaines d'entre-elles, situées dans une fourchette de notes (les copies notées entre 9 et 11/20 par exemple). La possibilité est alors laissée au correcteur de modifier les notations initialement codées. L'objectif est ici de voir l'impact éventuel de cette étape supplémentaire du processus de correction sur la fidélité de la mesure.

Références bibliographiques (format APA)

Gilles, J.-L., Detroz, P., Crahay, V. et al. (2011). La plateforme ExAMS, un "assessment management system" pour instrumenter la construction et la gestion qualité des évaluations des apprentissages. In Blais, Jean-Guy (Ed.) *Evaluation des apprentissages et technologie de l'information et de la communication - Tome 2* (pp. 11-40). Presses de l'Université Laval, Québec.

Detroz, P., (2016, avril). *L'évaluation des étudiants dans l'enseignement supérieur : vers un cycle qualité dans la réalisation de tests*. Communication présentée à l'Université du Luxembourg.

Scallon G. (2004). *L'évaluation des apprentissages dans une approche par compétences*. Bruxelles : De Boeck.

Babad, E.Y. (1980). Expectancy bias in scoring as a function of ability and ethnic labels. *Psychological Reports*, 46, 625-626.

Fajardo, D. M. (1985), Author Race, Essay Quality, and Reverse Discrimination. *Journal of Applied Social Psychology*, 15, 255–268.

Amigues, R., Bonniol, J.-J. & Caverni, J.-P. (1975). Les comportements d'évaluation dans les systèmes éducatifs : Influence d'une catégorisation ethnique sur la notation de productions scolaires. *International Journal of Psychology*, 10, 135–145.

AERA-APA-NCME. Standards for Educational and Psychological Testing (1999).

Bonniol, J., Piolat, M. (1971). Comparaison des effets d'ancrage obtenus dans une tâche d'évaluation. Expérience de multi-correction en mathématique et en anglais. In Actes du XVII^e Congrès international de psychologie appliquée, 8, 1179-1189.