

Vers un accompagnement docimologique instrumenté à l'aide la plateforme *e-C&QCST* pour les enseignants de l'Université de Liège qui pratiquent le testing standardisé

Auteurs

Crahay, Vincianne & Gilles Jean-Luc

Sommaire

La recherche communiquée par le SMART à l'AIPU Monastir en 2006 a montré que rectifier les épreuves a un impact significativement positif sur la validité et la fidélité de l'épreuve. En continuité avec cette étude, notre communication vise à faciliter la création de tests standardisés, plus fiables, plus efficaces, plus rapidement mis en œuvre, en vue de satisfaire au mieux les exigences actuelles du monde de l'enseignement/formation/ressources humaines. Notre objectif est d'améliorer l'accompagnement docimologique des enseignants de l'Université de Liège (ULg) à l'aide de la plate-forme électronique de construction et de gestion qualité de tests standardisés intitulée *e-C&QCST* (*electronic Construction & Quality Control in Standardized Testing*) réalisée à l'aide de logiciels libres et dans une philosophie « *Open Source* ».

Mots-clés

Evaluation, approche qualité, accompagnement, computer application

Texte

Dans le cadre du Congrès de l'AIPU Monastir 2006, notre recherche a analysé l'influence de l'accompagnement des enseignants par l'équipe du SMART-ULg (Système Méthodologique d'Aide à la Réalisation de Tests) sur les performances des étudiants et la qualité des épreuves. En d'autres termes, notre étude portait sur les effets de l'accompagnement sur la qualité des épreuves standardisées ainsi que sur les scores des étudiants. Cet accompagnement consiste en l'exposition et l'interprétation de tests statistiques (moyenne, écart type, alpha de Cronbach, *r*.bis,...) réalisés sur l'ensemble du test. En fonction de ces statistiques, l'enseignant peut mener des actions sur son évaluation pour rendre les résultats à son questionnaire à choix multiple plus fiables. Ces différentes actions sont : la valorisation d'un distracteur, le changement de réponse correcte, la suppression d'une question et le changement de tarif.

Les conclusions de notre communication se sont avérées favorables. En effet, rectifier les épreuves entraîne deux enjeux : au niveau de l'enseignant-évaluateur, la validité et la fidélité des tests s'améliorent. Autrement dit, après les rectifications, les épreuves mesurent mieux ce qu'elles sont sensées mesurer. Et au niveau des étudiants, globalement, dans le cadre des tests standardisés certificatifs il y a augmentation des moyennes des groupes et donc du nombre de réussites.

Depuis le milieu des années 90, l'accompagnement d'enseignants dans la réalisation des évaluations des acquis à l'Université de Liège est en constante augmentation. En effet, régulièrement de « nouveaux » enseignants font appel à nous pour avoir des conseils pédagogiques en terme de testing standardisé.

Dans ce contexte, nous proposons d'utiliser les résultats d'une recherche menée coordonnée par l'équipe du SMART¹ visant la mise en place de méthodes et d'outils de gestion de la qualité des épreuves standardisées en vue de satisfaire au mieux les exigences des enseignants

¹ Système Méthodologique d'Aide à la Réalisation de Tests, Unité de Soutien Logistique et de Recherche au sein de l'ULg.

et des étudiants. Il s'agit du projet *e-C&QCST*² qui a abouti à la réalisation d'une plate-forme électronique de construction et de gestion qualité de tests standardisés réalisée à l'aide de logiciels libres et dans une philosophie « *Open Source* » (Gilles & al, 2004, 2005).

La plate-forme *e-C&QCST* a pour but de formaliser et créer des procédures docimologiques qui permettent de maîtriser, quantifier et contrôler l'atteinte des objectifs qualité dans le cadre des étapes de construction d'une évaluation standardisée des compétences d'apprenants. Ensuite, ces procédures seront programmées. Les interfaces sont accessibles via l'Internet.

L'architecture repose sur huit modules intégrés en rapport avec le modèle « CGQTS », le cycle de Construction et de Gestion Qualité des Tests Standardisés utilisé par le SMART (Gilles et Leclercq, 1995; Gilles, 2002) :

1. l'analyse de l'enseignement en vue d'en dégager les points à évaluer ;
2. la mise en forme des épreuves (conception des modalités de questionnement) ;
3. la construction de tests en mode sécurisé (banques de questions et contrôle qualité *a priori*) ;
4. l'entraînement à l'aide de « tests à blanc » en ligne ;
5. l'aide à la mise en œuvre des tests certificatifs ou formatifs permettant entre autre la passation en ligne d'examens (dans des environnements surveillés) ou le couplage avec d'autres dispositifs d'acquisition automatique des réponses tels que la lecture optique de marques ;
6. la correction des épreuves avec contrôles qualité *a posteriori* et possibilités de rectifications ;
7. les feedbacks individualisés et personnalisés accessibles en ligne pour les apprenants ;
8. la macro-régulation avec récolte et analyse en ligne des avis des acteurs en vue d'améliorer le prochain test qui sera créé.

Par ailleurs, les contrôles qualité effectués lors de l'accompagnement des enseignants, se font en référence à une série de critères de qualité des évaluations standardisées proposés par le SMART (Gilles, 1998, 2002 ; Gilles et Lovinfosse 2004) :

- La validité : les scores des personnes évaluées doivent refléter ce que l'enseignant veut mesurer, permettre des inférences solides (validité de construct ou théorique) et couvrir les aspects importants des apprentissages (validité de contenu) ;
- La fidélité : une copie corrigée et classée dans une catégorie donnée doit bénéficier de la même mention si elle est corrigée par d'autres évaluateurs ou équipes d'évaluateurs (concordance inter juge) ou/et à un autre moment par le même correcteur ou la même équipe (consistance intra juge) ;
- La sensibilité : les mesures des compétences des évalués doivent être précises, refléter des phénomènes subtils ;
- La diagnosticité : les résultats des épreuves doivent permettre le diagnostic précis des difficultés d'apprentissage (et idéalement de leurs causes), des processus maîtrisés et de ceux qu'ine le sont pas ;
- La praticabilité : la faisabilité en termes de temps, de ressources en personnel et en matériel doit être assurée ;
- L'équité : tous les évalués doivent être traités de façon juste, en principe de la même manière ;
- La communicabilité : les informations non confidentielles relatives au déroulement du processus doivent être communiquées et comprises par les personnes engagées dans la passation du test ;
- L'authenticité : les questions doivent être en rapport avec des situations qui ont du sens pour les évalués, pertinentes par rapport au monde réel et propre à favoriser les transferts dans des contextes et des pratiques qu'on peut rencontrer dans la vie

Durant la construction de la plate-forme, des premiers essais d'accompagnement d'enseignants, futurs utilisateurs potentiels, ont permis de vérifier l'adéquation de la démarche

² *electronic Construction & Quality Control in Standardized Testing*

proposée par rapport aux besoins des acteurs, d'envisager des utilisations différentes de celles initialement prévues, de tester l'ergonomie logicielle au fur et à mesure de l'élaboration du produit.

La construction de la plate-forme touche à sa fin et nous commençons maintenant à l'utiliser à plein escient avec les enseignants de l'ULg demandeurs. L'hypothèse que nous émettons et qui sera développée durant cette communication concerne les débuts de sa mise en œuvre sur le terrain de notre université. Plus précisément, notre étude envisagera les avantages et les inconvénients liés à un accompagnement docimologique instrumenté via la plateforme e-C&QCST ainsi que l'impact de celle-ci, des procédures et contrôles qualité qu'elle propose, sur les pratiques des enseignants de l'Université de Liège en matière de construction de tests standardisés de qualité.

Références bibliographiques

- Crahay, V., Gilles, J.-L., Hausoul, E., Mainferme, R., Tinnirello, S., (2006), « Vers une qualité docimologique accrue des évaluations des acquis des apprenants à l'aide d'un cycle de construction et de gestion des tests standardisés ». *XXIII^{ème} Congrès de l'Association Internationale de Pédagogie Universitaire (AIPU^o, Innovation, formation et recherche en pédagogie universitaire*. Faculté de Pharmacie de Monastir, Tunisie
- De Finetti, B. (1965). Methods for discriminating levels of partial knowledge concerning a test item, *British Journal of Mathematical and Statistical Psychology*, 18, 87-123
- Gilles, J.-L. et Leclercq, D. (1995). Procédures d'évaluation adaptées à des grands groupes d'étudiants universitaires - Enjeux et solutions pratiquées à la FAPSE-ULG, in *Actes du Symposium International sur la Rénovation Didactique en Biologie*. Tunis : Université de Tunis.
- Gilles, J.-L. (1998). *Introduction d'un système qualité dans le processus de réalisation des épreuves certificatives de l'enseignement supérieur*. Communication à la demi-journée de sensibilisation à la Qualité – Université de Liège, 5 octobre 1998.
- Gilles, J.-L. (2002). Qualité spectrale des tests standardisés universitaires, thèse de doctorat en Sciences de l'Education. Liège : Université de Liège, Faculté de Psychologie et des Sciences de l'Education
- Gilles, J.-L. et Lovinfosse, V. (2004). Utilisation du cycle SMART de gestion qualité des évaluations standardisées dans le contexte d'une Haute Ecole : regard critique en termes de validité, fidélité, sensibilité des mesures, diagnosticité, praticabilité, équité, communicabilité et authenticité, Communication au *XIV^{ème} congrès de l'Association Mondiale des Sciences de l'Education (AMSE)*. Santiago du Chili, 10-14 mai 2004.
- Gilles, J.-L. (2004). Feedbacks diagnostiques via internet : un système individualisé de communication des niveaux de performances des étudiants dans le cadre d'évaluations standardisées, in *Actes du 3^{ème} Congrès des chercheurs en éducation « (Re)trouver le plaisir d'enseigner et d'apprendre – Construire savoirs et compétences »*, Bruxelles : Administration générale de l'Enseignement et de la recherche scientifique
- Gilles, J.-L., Piette, S.-A., Detroz, P. et Pirson, M. (2004). Le projet de plate-forme électronique de construction et de gestion qualité de tests standardisés *e-C&QCST*, Communication au *XIV^{ème} congrès de l'Association Mondiale des Sciences de l'Education (AMSE)*. Santiago du Chili, 10-14 mai 2004.
- Gilles, J.-L., Piette, S.-A., Detroz, P., Tinnirello, S., Pirson, M., Dabo, M. et Lê, H. (2005), « The electronic Construction and Quality Control in Standardized Testing platform project (e-C&QCST) », in Demetriou, A., Dochy, F. & al (Eds), *European Association For Research on Learning and Instruction (EARLI) – 11th Biennial Conference of EARLI: Book of Abstract*. Nicosia: University of Cyprus, Department of Educational Sciences
- Leclercq, D. (1982). Confidence marking, its use in testing, in B. Choppin & N. Postlethwaite (eds.), *Evaluation in Education : International Review Series*, Oxford: Pergamon, vol. 6, n^o2, pp. 161-287
- Leclercq, D. & Gilles, J.-L. (1993). Hypermedia : Teaching Through Assessment. In D. Leclercq et J. Bruno (Eds), *NATO ASI Series F, Item Banking: Interactive Testing and Self Assessment*. Berlin: Springer Verlag, Vol. 112, pp. 31-47
- Leclercq, D., Peeters, R., Reggers, T., Charlier, J.E., De Ketele, J.M., Delhaxhe, M., Dupont, P., Lambert, J.P., Lambotte, J.P., Loeckx, E., Pilatte, A., Romainville, M. & Wolfs, J.-L. (1997). *Franchir le cap des candidatures – Inscriptions et réussites en candidatures universitaires dans la Communauté Française de Belgique de 1988 à 1995*. Bruxelles : Conseil Interuniversitaire de la Communauté Française
- Jans, V. & Leclercq, D. (1999). Mesurer l'effet de l'apprentissage à l'aide de l'analyse spectrale des performances, in C. Depover & B. Noel (Ed.), *Evaluation des compétences et des processus cognitifs*. Bruxelles : De Boeck, pp. 303-317
- Reach, G., Zerrouki, A., Leclercq, D. & d'Ivernois, J.-F. (2005). Adjusting insulin doses: from knowledge to decision, *Patient Education and Counseling*, Vol. 56, Issue 1, pp. 98-103
- Shufford, A., Albert, A. & Massengil, N.E. (1966). Admissible probability measurement procedures, *Psychometrika*, 31, 125-145