

Pourquoi les enseignants travaillant avec le SMART (sous ?)-exploitent-ils les potentialités du système de feedbacks aux étudiants ?

Vinciane Crahay, Aurélie Piazza, Laurent Leduc, Pascal Detroz

Introduction

En un peu moins de 20 ans, le concept de réflexivité (Schön, 1994) a acquis une telle portée heuristique, qu'il est maintenant considéré comme étant l'un des moteurs principaux du développement professionnel des enseignants (Charlier, 2000).

La mise en application de ce concept offre également un matériau riche et utile au chercheur qui essaie de comprendre le fonctionnement cognitif des enseignants. En effet, comme le souligne Perrenoud (2001), la réflexivité implique le fait de *penser l'action professionnelle au plus près, y compris dans ses composantes les moins visibles : la réflexion, avant, pendant et après l'action (p. 45)*. En ce sens, elle rend cette action plus visible, elle en améliore la lecture.

Même si les domaines de la pédagogie et ceux de l'organisation s'échangent les concepts sans trop de difficultés (le concept de compétence en est un bel exemple), et malgré le fait que la notion d'organisation apprenante (Argyris et al., 1978) ait aujourd'hui rejoint le sens commun, on parle assez peu de réflexivité institutionnelle.

Or, la pratique réflexive, non plus abordée en termes individualisants, mais bien portée au niveau d'une organisation, nous semble porteuse d'intérêt. Pour l'organisation, certes, qui y trouve une opportunité d'évolution, mais également pour la communauté, notamment scientifique, qui peut ainsi mieux comprendre les rouages de celle-ci.

C'est dans cette perspective que nous avons écrit ce document. Nous proposons d'y présenter une réflexion, praxéologique et réflexive, sur les actions du Système d'Aide à la Réalisation de Tests (SMART) de l'Université de Liège. Celle-ci nous semble intéressante à plus d'un titre.

D'une part, la nature même du SMART nous semble mériter une certaine attention. En effet, celui-ci peut se définir comme un centre docimologique dont la vocation essentielle est d'aider les enseignants de l'ULg à concevoir, administrer et piloter les évaluations qu'ils proposent à leurs étudiants. A notre connaissance, une telle structure est peu fréquente à l'Université, le conseil docimologique étant le plus souvent le parent pauvre du conseil pédagogique dans lequel il se retrouve souvent noyé.

D'autre part, le SMART est une organisation qui fonctionne depuis plus d'une quinzaine d'années. Cela nous donne un certain recul par rapport à la manière d'aborder notre métier de conseiller docimologique. Cela nous permet également d'avoir un avis critique sur les évolutions des pratiques d'évaluation au sein de notre établissement et, sans doute, au sein des établissements universitaires en général.

Ajoutons également que nous avons nourri, dès l'origine, l'ambition de recueillir quelques indicateurs, certes frustes mais bien réels, de notre fonctionnement. Selon nous, ceux-ci donnent incontestablement du corps à notre analyse.

Ce papier est donc un peu hybride. Entre présentation de dispositif, retour d'expérience et analyse de cas, il a pour ambition, sur base d'une analyse réflexive de 15 ans d'actions, de dégager les lignes de forces et de faiblesses de notre accompagnement docimologique. Il a aussi pour ambition d'éclairer une face souvent dans l'ombre de l'activité des enseignants universitaires, à savoir les modalités d'évaluation (Romainville, 2006).

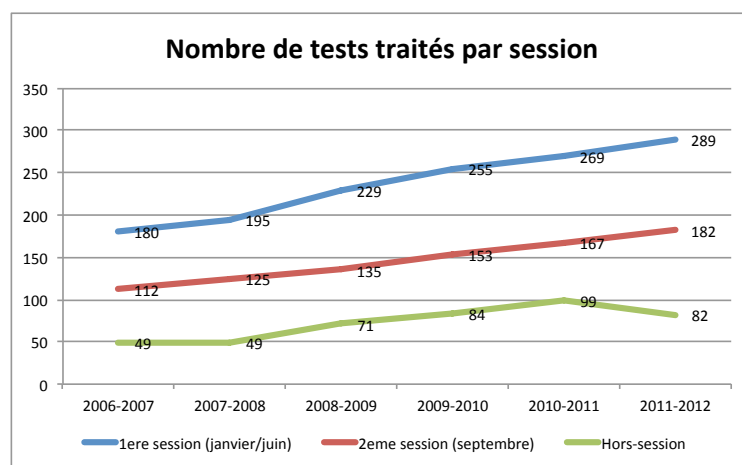
Le Système Méthodologique d'Aide à la Réalisation de Tests (SMART)

Porté sur les fonds baptismaux par D. Leclercq et J-L Gilles à la fin des années nonante, l'ancêtre du SMART, le CAFEIM, est une structure facultaire située au sein de la Faculté de Psychologie et des Sciences de l'Éducation (FAPSE) de l'ULg. En conformité avec les travaux scientifiques de D. Leclercq (1986), ce service s'oriente assez naturellement vers le conseil méthodologique et logistique dans le domaine du Testing Standardisé et, plus particulièrement, dans celui des questions à choix multiples. Peu à peu, les clients internes, essentiellement constitués de professeurs de la FAPSE, se diversifient et au début des années 2000, de nombreux enseignants, issus de toutes les facultés, font appel aux services du CAFEIM. Afin d'éviter toute une ingénierie de facturation interne en lien avec ce soutien docimologique, Monsieur le Recteur W. Legros, décide de subsidier directement cette équipe et de lui donner une vocation clairement interfacultaire : le SMART est né.

Depuis lors, le volume d'activité de cette équipe n'a pas cessé de croître. Actuellement, le SMART compte environ 5 ETP spécifiquement dédiés à l'accompagnement docimologique d'évaluations certificatives et formatives. En 2012, nous avons accompagné environ 550 épreuves.

Il est l'heure pour nous de tirer quelques tendances. Dans le cadre de cet article, nous analyserons les raisons en lien avec une augmentation de notre volume d'activité.

Une augmentation linéaire du volume de nos activités

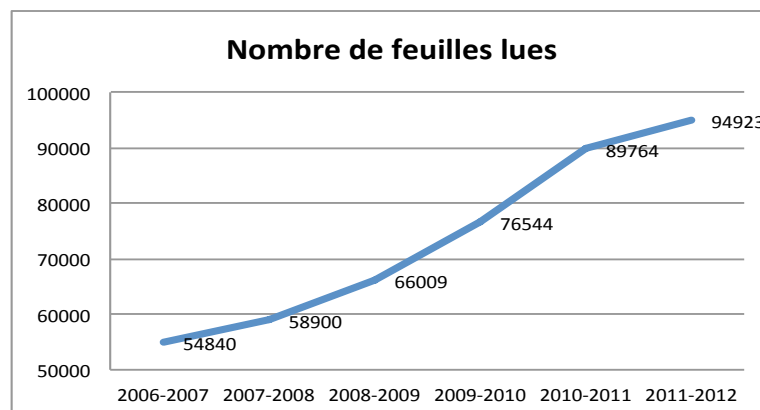


Graphique 1

Comme le montre le graphique 1, le nombre des tests standardisés a fortement évolué lors de ces dernières années. Aujourd'hui, nous accompagnons pas moins de 553 tests

par an. La toute grande majorité de ceux-ci utilisent le QCM comme modalité de questionnement. Cela s'explique par des raisons pratiques. Les enseignants, aux premiers abords, considèrent que la plus-value du SMART est, avant toutes choses, logistique. Dépassés par la massification du nombre d'étudiants et quasi apeurés devant une tâche d'évaluation jugée colossale, ils sont très heureux d'automatiser la fastidieuse étape de correction de leur QCM. Or, justement, notre aide logistique leur fait surtout gagner un précieux temps lorsqu'ils utilisent cette modalité de questionnement. Nous offrons également une semi-automatisation de correction de productions écrites (Gilles, Detroz, et al. 2011), mais la plus-value de cette plateforme de correction est avant tout méthodologique : elle ne diminue pas le temps nécessaire à la correction et est dès lors moins souvent utilisée.

Si le nombre de tests augmente, c'est également le cas, et de manière plus flagrante encore, du nombre de feuilles corrigées (voir graphique 2) qui passe de 54840 copies corrigées en 2006-2007 à 94923 copies en 2011-2012. Autre statistique, 55% des étudiants de l'ULg, en 2012, ont au-moins eu un test standardisé corrigé par le SMART.



Graphique 2

Notons par ailleurs que ces augmentations sont également dues à une explosion de nos clients internes. Alors qu'en 2006 nous travaillions directement avec 86 enseignants de notre université, nous avons accompagné 142 enseignants en 2012.

Nous nous trouvons donc devant ce qu'il faut bien appeler une explosion des modalités d'évaluation standardisée au sein de notre institution.

Cette évolution peut sembler a priori étonnante. En effet, aux Etats-Unis comme en Europe, le testing standardisé n'a pas bonne presse. Certes, certains le défendent (Phelps, 2005) et il présente un certain nombre d'avantages bien décrits dans la littérature. Par exemple, en synthétisant les travaux de Nitko (2004), Leclercq (2005) et Linn & Gronlund (2000) nous concèderons au QCM leur capacité à évaluer rapidement une matière étendue, une grande équité dans la correction (puisque le scoring est automatisé) et une certaine force diagnostique (en ce sens qu'elle autorise des feedbacks rapides et précis).

Mais cette modalité d'évaluation présente aussi de nombreux désavantages qui semblent difficilement conciliables avec certaines conceptions actuellement dominantes

de la formation universitaire. Elles ne permettent pas ou rendent difficile l'évaluation de performances complexes et ne prennent pas en compte les habiletés et le savoir-être.

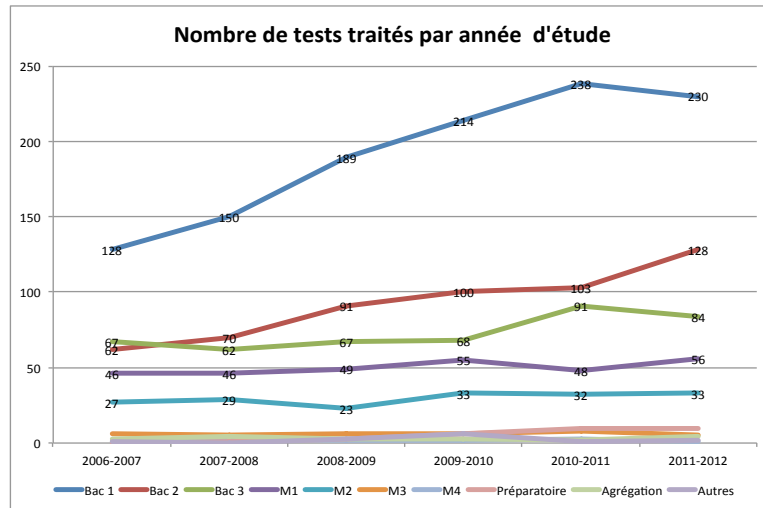
Pire, elles semblent en totale inadéquation avec une approche par compétences, voire avec une approche programme qui semble de plus en plus s'ériger en dogme en pédagogie universitaire. Ainsi, même si de nombreux auteurs ayant réfléchi à la question peinent encore à définir ce que doit être l'évaluation dans une approche par compétences, ils sont au-moins d'accord sur un point : les règles docimologiques traditionnelles ne s'appliquent plus, ou s'appliquent mal. Ainsi, De Ketele et Gerard (2005) concluent : « *Peut-on espérer un jour disposer d'une éduométrie satisfaisante pour évaluer des compétences complexes ? Rien n'est moins sûr dans l'état actuel de nos connaissances* », Scallon (2004) précise que « *La démarche à implémenter [pour évaluer les compétences] demeure inédite* », et Tardif (2006) écrit que « *Les exigences de l'évaluation des compétences sont nombreuses et elles posent des défis de taille. D'aucun pourraient dès maintenant penser qu'une telle entreprise est impossible étant donné les théories et les instruments développés à ce jour dans le domaine de l'évaluation des apprentissages. (p.134)* ».

Si le chemin reste long et sinueux (Scallon, 2004), une chose semble évidente : le testing standardisé est inapte à évaluer des compétences.

Quels sont dès lors, dans notre université où l'approche par compétences percole doucement dans les pratiques pédagogiques, les éléments qui expliquent cette progression fulgurante des QCM ?

- 1) La massification des étudiants. De nombreux nouveaux utilisateurs de nos services nous contactent avec un discours du type : « *je n'aime pas les QCM, mais le nombre d'étudiants dans la filière dans laquelle j'enseigne vient d'augmenter de façon significative et je n'ai aujourd'hui plus le choix* ».

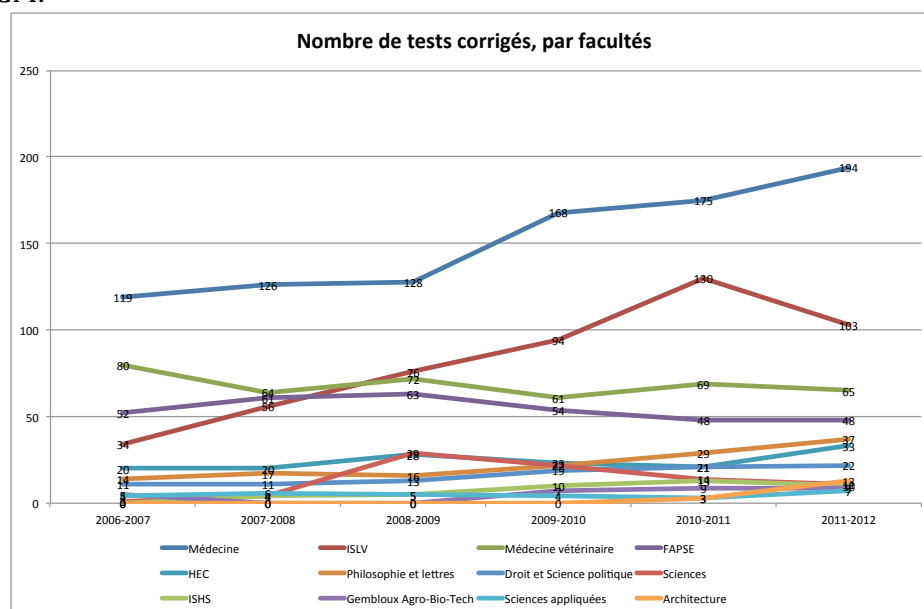
Cette analyse, basée sur l'observation, est également soutenue par des données plus empiriques. Si l'on se concentre sur le graphique 1, nous constatons que les tests standardisés sont plus fréquents en première session qu'en seconde session (qui, par définition, compte moins d'étudiants). Le graphique 3 apporte également de l'eau à ce moulin.



Graphique 3

Comme on le voit, les années d'études qui drainent le plus d'étudiants sont aussi, de manière flagrante, celles qui ont le plus recours à nos services. Evitons toutefois un discours abusif: il est évident que la maîtrise des catégories de performance pour lesquelles l'utilisation de QCM est appropriée, sont aussi celles que vise le plus souvent les enseignements du premier degré. C'est le cas par exemple de la connaissance, de la compréhension ou de l'application.

Une analyse filière par filière est également éclairante à ce sujet. Comme le graphique 4 l'exprime clairement, notre plus gros client interne est la faculté de médecine qui est aussi celle qui draine les plus grands auditoires en premier bac. Mais sur ce graphique, nous constatons aussi que le droit et les HEC, dans lesquels nous retrouvons respectivement les deuxième et troisième auditoires en termes de nombre d'étudiants, recourent relativement peu, proportionnellement, aux QCM.



Graphique 4

Au vu de ces données, nous pourrions faire l'hypothèse qu'il existe des raisons épistémologiques ou disciplinaires de ne pas trop recourir aux QCM. Ainsi, par

exemple, on pourrait expliquer que le droit est une discipline ou l'argumentation prend une place tellement importante qu'il n'est pas envisageable de standardiser le testing. Ou que les études de sciences économiques et de gestion impliquent le développement et l'évaluation de soft skills qui ne peuvent s'évaluer qu'à l'oral. Cette hypothèse est plausible et mériterait d'être creusée plus avant.

Elle n'explique toutefois pas les disparités entre Gembloux Agro-Bio-Tech et la faculté de Sciences Appliquées qui comptent des auditoriums de tailles quasi similaires et forment tout deux des ingénieurs mais qui recourent de manière très différente à nos services. Peut-être faut-il chercher les raisons de cet écart dans des normes facultaires ou dans l'image que se font les facultés de l'efficacité du SMART.

2) La vision docimologique du SMART.

Le SMART a développé, au fil du temps, une posture praxéologique vis-à-vis des enseignants qu'il accompagne. Notre maître-mot est « co-construction des modalités d'évaluation ». Evidemment, nous sommes conscients des limites du testing standardisé. De ses lignes de forces aussi. Nous avons également développé une méthodologie, un cycle qualité en 8 étapes (Gilles, 2002) qui permet à nos partenaires, avec l'appui de notre accompagnement, de tendre directement vers plus de qualité (voir fig. 1).

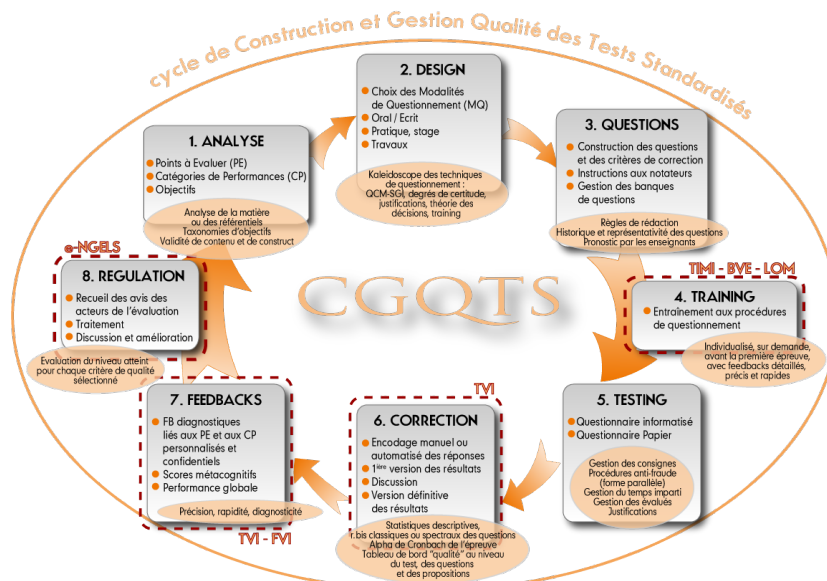


Figure 1

Nous savons donc comment nos enseignants devraient s'y prendre pour améliorer la qualité de leurs évaluations et avons développé les outils pour les y aider (par exemple le logiciel ExAMS - Gilles, Detroz, et al. 2008 - ou le logiciel TQROL - Detroz, Crahay, Piazza, 2013). Mais nous sommes également profondément respectueux des contraintes dans lesquelles naviguent nos partenaires. En tension entre les activités de recherche et d'enseignement, en tension entre les activités d'enseignement et les examens, rares sont ceux qui, de prime abord, sont prêts à consacrer des moyens importants à améliorer leur dispositif d'évaluation. Nous l'avons déjà mis en exergue, la plupart des

enseignants viennent d'abord chercher une aide logistique, parfois une aide à caractère urgent et salvateur, et c'est avec plaisir que nous la leur offrons.

A côté de cela, nous ne considérerions pas notre mission comme réussie si nous n'arrivions pas à enclencher, consécutivement à cette aide logistique, des démarches méthodologiques de qualité.

Ainsi, nous avons œuvré à fournir de manière systématique et automatisée un indicateur de la qualité de leurs QCM : la corrélation point bisériale, le rpbis. Cet indicateur, dont vous trouverez (entre autres) la description sur le site du SMART à www.smart.ulg.ac.be/smartweb/outils/feedback/statsGroupe_rpbis.pdf l'adresse permet de s'assurer que ce sont bien les meilleurs étudiants, en tendance, au test qui répondent correctement à une question donnée et, en corollaire, que ce sont bien les étudiants, en tendance, les moins bons au test qui y échouent. Cet indicateur est fourni et analysé par nos soins pour l'ensemble des 550 tests que nous corrigeons par an.

L'intérêt du rpbis est de relever les questions pour lesquelles le pattern ci-dessous n'est pas rencontré. Par exemple, lorsque les meilleurs étudiants au test choisissent massivement un distracteur qui, par définition, n'est pas la réponse jugée comme étant la bonne par l'évaluateur. Cela peut-être le cas quand l'enseignant s'est lui-même trompé en nous communiquant la réponse correcte, mais la situation est parfois plus ambiguë. Par exemple, avec cet éclairage statistique, il peut se rendre compte qu'il existe un autre distracteur qui s'avère lui aussi être une réponse acceptable. Ce cas est fréquent et amène l'enseignant à mettre en place une stratégie correctrice. Par exemple, il peut modifier la réponse correcte ou valoriser un second distracteur. Il peut aussi supprimer la question ou la valoriser pour l'ensemble des répondants. Comme le montre le tableau 1, l'application de ces stratégies correctrices est fréquente.

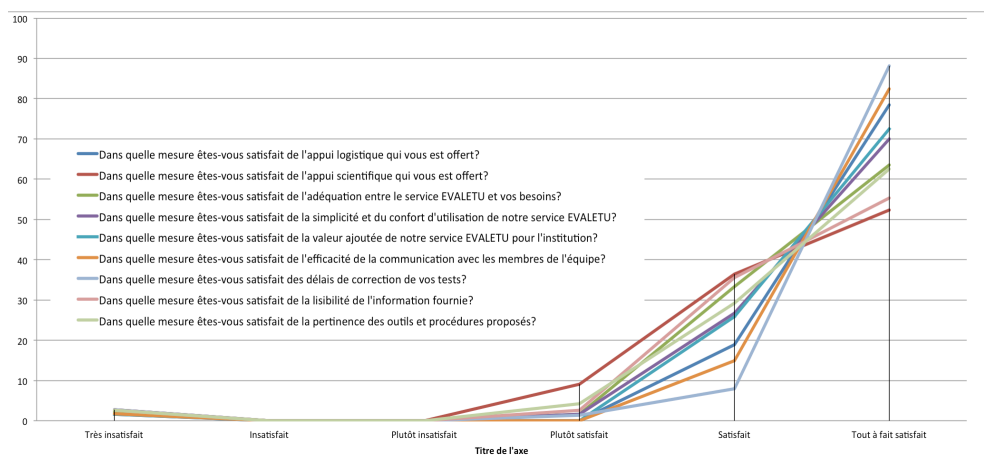
Sur l'ensemble des 492 tests corrigés en 2009-2010, 206 (soit 42 %) ont été modifiés suite à l'analyse des rpbis	
Suppression d'une question	94 tests
Valorisation d'un distracteur	52 tests
Changement de réponse correcte	60 tests

Outre l'effet direct sur la qualité de la mesure en évaluation (en l'absence d'analyse rpbis, 60 tests sur les 492 - soit 12 % - comprenaient une erreur quant à la réponse correcte indiquée par l'enseignant pour au-moins une question) il s'agit aussi d'un premier pas vers une démarche qualité en évaluation. Ce premier pas a l'avantage d'être très pragmatique (le professeur voit tout de suite son intérêt), pas du tout contraignant pour l'enseignant (puisque l'analyse statistique est automatique et son interprétation faite par le SMART) et très respectueux de l'écologie de l'enseignant puisque lui seul prend les décisions par rapport à son test.

Ce premier pas est néanmoins fondamental dans notre approche. Il permet de faire prendre conscience à l'enseignant des dangers de l'improvisation et de

l'approximation en évaluation. Il nous permet aussi d'atteindre un statut différent. L'enseignant se rend compte qu'au delà de l'aide logistique, nous pouvons lui offrir une aide plus méthodologique. La relation de confiance s'installe. Il considère, quelque part, que nous lui avons sauvé la mise. Il est prêt à nous faire confiance dans le cadre d'un accompagnement méthodologique. Bien sur, et nous avons beaucoup de compréhension pour cela, tous n'enclenchent pas tout de suite des démarches qualité. Celles-ci sont contraignantes, notamment en temps, et nous avons déjà indiqué l'empathie dont nous faisons preuve face à la charge de travail des enseignants. Mais pour la plupart de ceux que nous avons accompagné scientifiquement, ce premier pas a agi comme un déclencheur, voire comme un révélateur.

Nous sommes convaincus que cette vision souple et assumée de l'accompagnement, ne mettant pas d'emblée des considérations théoriques complexes en avant, mais les réservant à des moments où la relation collaborative a été rendue possible, a eu un impact très important sur l'image institutionnelle que nous dégageons. Notons également que lorsque nous avons interrogé les 150 enseignants avec lesquels nous collaborons (Taux de réponse, 51 %), quasi tous se sont montrés au-moins satisfaits de notre accompagnement scientifique (voir graphique 5).



Graphique 5

3) Notre approche qualité.

Dès le début de nos activités, nous avons mis en place des démarches qualité performantes. Ainsi, la traçabilité de chaque feuille est assurée et, en cas de litige, nous pouvons en quelques minutes retrouver la feuille de l'étudiant et, s'il y en a eu, justifier toutes les opérations réalisées sur cette même feuille. Par exemple, si l'étudiant a oublié de noter son matricule, mais qu'il a écrit son nom, notre procédure nous le signale et nous pouvons rechercher et compléter ses coordonnées à l'aide de l'annuaire de l'ULg. Cette opération sera consignée dans la fiche de suivi du test, en toute transparence. Ainsi, l'ensemble des opérations nécessaires (schématisé fig. 2) donne lieu à pas moins de plusieurs dizaines de contrôles qualité assurant ainsi le caractère robuste de nos résultats. Pour arriver à une telle maîtrise, nous avons dû notamment concevoir le logiciel TVI, qui est au cœur de notre processus de traitement et qui est aujourd'hui utilisé à l'Université d'Ho chi min au Vietnam et à l'Université Simon I Patino en Bolivie.

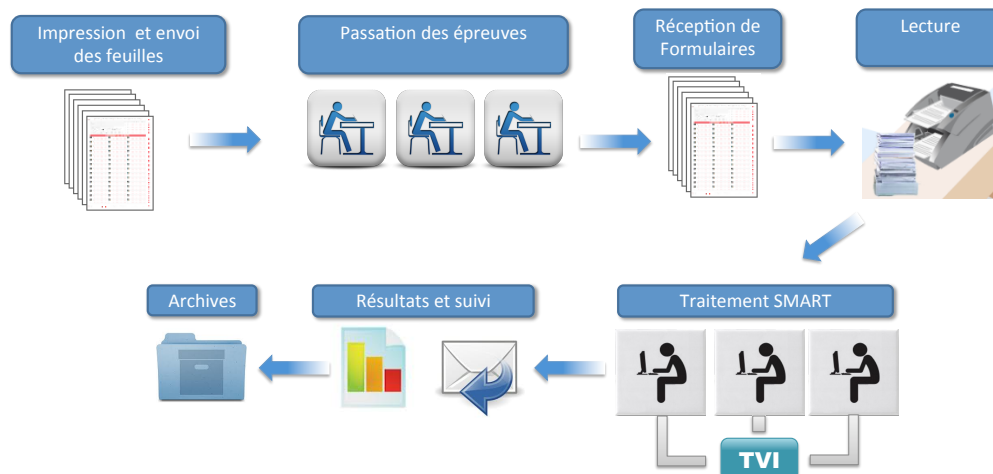


Figure 2

Les conséquences de cette approche qualité sont extrêmement positives. Jamais nous n'avons produit de résultats incorrects. A l'heure où le recours au Conseil d'Etat de la part d'étudiants insatisfaits est entré dans les normes universitaires, jamais nous n'avons été pris en défaut. Cela assure une image positive dans l'institution, ce qui peut également insuffler la confiance nécessaire à la collaboration avec nos collègues.

Si l'on adopte une perspective historique face à ces contrôles qualité, on perçoit une évolution vers plus de professionnalisme. Cela a été particulièrement marquant durant les cinq dernières années, en ce qui concerne la distribution des feedbacks aux étudiants. Plus récemment également œuvré à une meilleure traçabilité de nos actions de corrections

Le feedback aux étudiants

Scientifiquement, dans la lignée des travaux de Nicol et MacFarlane-Dick (2006) et de Schute (2007), nous avons toujours eu conscience de la nécessité d'offrir aux étudiants un feedback rapide, diagnostique et personnalisé. L'étape feedback est d'ailleurs la septième du cycle qualité présenté en figure 1. Jusqu'il y a quelques années, les étudiants se connectaient sur notre site, choisissaient le cours, encodaient leur matricule et leurs initiales et ils avaient accès à un tel feedback. Le système a ainsi fonctionné pendant quelques années à la satisfaction de tous.

Nous avons toutefois pris conscience que les données ainsi fournies étaient assez peu protégées, le matricule des étudiants étant une information que l'on peut assez aisément obtenir. Dans un contexte sociétal où la tension par rapport à la confidentialité des données personnelles ne cesse d'augmenter, il nous a dès lors semblé nécessaire de mieux protéger ce type d'information. Cela a donné lieu à un projet commun avec le Service Général d'Informatique de notre institution qui a permis de faire en sorte qu'aujourd'hui, les étudiants qui souhaitent obtenir leurs feedbacks personnels doivent se connecter au site institutionnel de l'ULg avec leur identifiant et leur mot de passe pour ensuite avoir accès à un espace personnel d'où ils pourront obtenir les rétroactions.

Depuis la mise en production de ce nouveau type de distribution des feedbacks (juin 2012), plus de 100.000 rétroactions ont été demandées par les étudiants.

La raison d'un tel engouement peut en partie s'expliquer par différents facteurs :

- a. Le professeur peut lier le questionnaire au feedback. Ainsi, en plus de voir si telle ou telle question est réussie ou ratée, l'étudiant peut remettre les numéros en relation avec l'énoncé qui est disponible sur le questionnaire annexé. Il obtient donc le corrigé de **son** test. De plus, le fait de lier des rubriques matières, des chapitres mais aussi des processus mentaux (comme la taxonomie de Bloom par exemple) aux différentes questions permet également aux étudiants d'avoir une vue d'ensemble, mais également une vue analytique de ses performances et, ainsi, de pouvoir adapter son étude et réguler son apprentissage.

Les possibilités ainsi offertes aux enseignants sont en phase avec les prescrits de la littérature. En effet, d'après Kulhavy et Stock (1989), un feedback efficace implique de mettre à disposition de l'étudiant deux éléments d'information : la *vérification* - « aussi appelée connaissance des résultats, elle informe l'apprenant de l'exactitude de ses réponses », communiquant ainsi le « juste/pas juste ou le pourcentage général de réponses correctes » (Shute, 2008, pp. 8 - 10) - et l'*élaboration* - « aspect informatif du message, procurant des indices permettant de guider l'étudiant vers la réponse correcte » (Shute, 2008, p. 8) -. Or, au regard des options décrites ci-dessus, les services offerts par le SMART permettent de délivrer des feedbacks incluant des informations relevant de ces deux catégories.

En effet, en invitant les enseignants à reconnecter les feedbacks aux énoncés des questions d'origine, mais aussi et surtout à cibler à travers celles-ci des niveaux taxonomiques d'objectifs ou des points de matières associés, le dispositif mis en place par le SMART dépasse la dimension strictement vérificatoire pour investir celle de l'élaboration, d'abord à travers deux de ses variations telles que décrites par Shute (2007, p. 8) : « elaboration can a) address the topic, b) address the response ». Au-delà, l'occasion donnée en particulier aux enseignants de mettre en lumière les relations entre énoncés de questions, feedbacks et *objectifs pédagogiques* rejoint les recommandations de divers experts des critères de qualité du feedbacks, parmi lesquels Nicol et MacFarlane-Dick - « where feedback is given it is important that it is related to (and that students understand its relation to) goals, standards or criteria » (2007, p. 9) - ou encore Kluger et DeNisi - « When the feedback sign is negative, people choose to increase their effort (rather than lower the standard) when the goal is clear,... » (1996).

Bien que supportée par des éléments théorique, ces options ne sont pas tout le temps utilisée par les professeurs, qui ne donnent alors aux étudiants que leur note globale à l'examen.

- b. Cet engouement peut aussi s'expliquer par le fait que, lorsqu'un nouveau feedback est disponible, une pastille s'affiche dans le bureau virtuel de l'étudiant. De la sorte, à côté des icônes référant à ses mails, aux petites annonces, aux documents disponibles, se trouve une petite icône SMART indiquant le nombre de nouveau FB disponibles. Il n'est donc plus question ici de devoir faire la démarche de se connecter à notre site et d'encoder les informations nécessaires à la récupération du document (parfois mal connues par l'étudiant, comme son matricule) : le feedback vient à eux et ils sont dans tous les cas informés lorsqu'un nouveau feedback est disponible.

La disponibilité immédiate du feedback par l'étudiant sur son bureau virtuel sans détours par de fastidieuses manipulations techniques ou recherches en ligne contribue à une forme de simplification du message et de son caractère - littéralement - *accessible*. Le bénéfice de cette accessibilité procède de mécanismes à rapprocher selon nous de ceux du respect des critères de *concision* - « if feedback is too long or too complicated, many learners will simply not pay attention to it, rendering it useless. Lengthy feedback can also diffuse or dilute the message » (Shute, 2008, p. 9) - et de *lisibilité* du feedback de qualité - un feedback de qualité peut être délivré en s'assurant de la lisibilité du feedback et en évitant si possible l'écriture manuscrite [ex. utiliser la voie électronique, par mail ou par un forum.

- c. Lorsque les degrés de certitude (DC) sont utilisés dans une évaluation, l'étudiant peut obtenir, en plus de ses notes à l'examen, son niveau de réalisme, de métacognition. En effet, nous fournissons plusieurs indicateurs : le taux d'exactitude (pourcentage moyen de réponses correctes), la certitude moyenne utilisée ainsi qu'un graphique de réalisme indiquant si l'étudiant se surestime (lorsqu'il indique un haut DC pour une réponse incorrecte) ou se sous-estime (lorsqu'il indique un faible DC alors que la réponse indiquée est correcte). En effet, il peut arriver qu'un étudiant ait toutes les réponses correctes à un test mais ayant utilisé les DC 0 à toutes les questions, il n'obtiendra finalement qu'une note de 13/20. Il est donc primordial que les étudiants puissent avoir une rétroaction quant à leur utilisation des degrés de certitude.

Ces trois grands facteurs mériteraient d'être généralisés afin que le feedback soit le plus informatif et utile aux étudiants. De la sorte, les étudiants auraient un maximum d'information contenue dans un seul et même document.

Cependant, même si nous prônons l'utilisation de ces aides à la régulation, tous les professeurs ne se montrent pas preneurs d'une telle mise à disposition de l'information.

En effet, certains avancent notamment les raisons suivantes à la non publication des feedbacks. Tout d'abord, ils ne souhaitent pas joindre le questionnaire afin de pouvoir réutiliser les questions ultérieurement. Ils n'ont parfois pas non plus l'énergie pour relier les questions aux chapitres et aux catégories de performance. Nous pouvons comprendre ces raisons, mais elles entraînent, de facto un appauvrissement du feedback : l'étudiant voit simplement qu'il a

répondu telle proposition à la question X mais ne sait pas à quoi cela fait référence.

D'autres professeurs avancent quant à eux que la note de cet examen fait partie d'un ensemble plus vaste du processus évaluatif (questions ouvertes, partie orale, TP,...) et que la cote qui apparaîtrait sur leur feedback ne serait pas celle indiquée sur leur bulletin. Ils préfèrent alors éviter de publier la rétroaction afin d'éviter quantité de mails de réclamation.

La traçabilité de nos corrections

Citons également une autre évolution de nos contrôles qualité. Si l'on reprend l'exemple de l'étudiant qui a oublié de noter son matricule sur sa feuille, nous avons mentionné qu'un opérateur humain complétait son formulaire. Cela se fait jusqu'à ce jour en notant directement celui-ci sur la feuille. Si, pour le matricule, cela ne semble pas poser de problème majeur, les interventions directement sur la copie de l'étudiant nous semblent délicates à effectuer. C'est le cas par exemple lorsque l'étudiant a répondu de manière ambiguë sur son formulaire (par exemple en noircissant 2 coches, mais en indiquant par une marque laquelle doit être prise en compte). Cela n'est pas encore arrivé, mais techniquement, nous pourrions comprendre que l'étudiant s'étonne d'une manipulation directe de sa feuille. C'est la raison pour laquelle nous utiliserons dès janvier d'autres procédures où les modifications ne se feront plus sur la feuille de l'étudiant, mais s'appuieront sur une copie scannée de sa feuille. Ainsi nous assurons une traçabilité complète des éléments manipulés sur la copie virtuelle.

Ces évolutions nous semblent nécessaires. Le niveau d'exigence des étudiants par rapport à la qualité des corrections de leur copie croit de manière très nette et nous devons anticiper ces nouveaux impératifs.

4) En conséquence, une image institutionnelle positive.

Comme vous le constaterez à la lecture du graphique ci-joint notre image institutionnelle est résolument positive. Les raisons de cet état de fait résident sans doute dans les éléments décrits ci-dessus. Permettons-nous, à la fin de cet article, d'insister sur notre posture praxéologique. Comme vous le constaterez à la figure 3, nous plaçons les besoins de nos partenaires au centre de nos préoccupations.

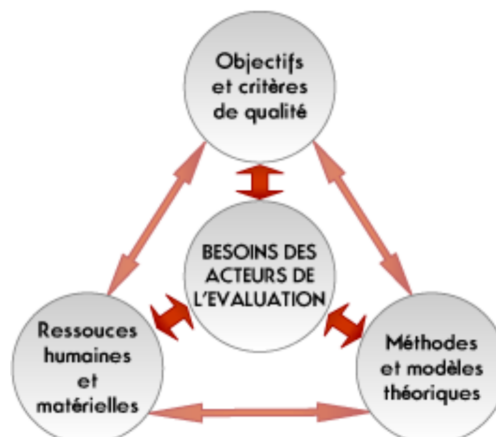


Figure 3

Il serait simple, pour nous, d'appliquer de manière machinale les préceptes de la docimologie et de guider de manière automatique nos collègues à travers nos outils. Mais nous sommes convaincus qu'être trop prescriptifs provoquerait un rejet instantané de nos services. Nous sommes donc amenés à jouer les équilibristes et proposer au cas par cas, après une analyse fine de la situation, les procédures et les outils les mieux à même de répondre à ces besoins. Cela nécessite quelques compétences particulières, comme par exemple une capacité de communication et de conviction, mais également, une capacité de renoncement. Il est parfois frustrant de savoir ce qu'il faut faire mais de constater que cela ne se fera pas, le plus souvent faute de temps disponible chez nos collègues. Mais une attitude compréhensive et bienveillante nous permet de garder le contact avec leurs préoccupations. Cela nous impose parfois de revoir certaines ambitions à la baisse. Il est important, selon nous, d'aborder la collaboration dans une perspective longitudinale et de considérer l'amélioration de la qualité docimologique comme un processus évolutif, de devenir en quelque sorte un adepte de la théorie des petits pas qui, mis bout à bout, amèneront à cette qualité. Vous l'aurez compris, patience, adaptabilité, compréhension, conviction et bienveillance, en plus de la maîtrise technique, sont des qualités nécessaires à ce type d'accompagnement

Conclusion

Nous voulions, à travers ces quelques lignes rendre compte de notre expertise longue de 15 ans dans l'accompagnement docimologique. Nous espérons que le lecteur aura pu y trouver une information susceptible de nourrir sa réflexion.

Les défis, pour notre équipe, restent nombreux. Si cet article rend compte de notre capacité à accompagner nos partenaires dans le cadre du testing standardisé, il reste muet sur notre capacité à les fédérer autour d'une amélioration de l'ensemble de leur dispositif d'évaluation. Nous l'avons mentionné, nos outils sont prêts pour les accompagner dans les évaluations de performances complexes, mais ces outils, contrairement à ceux que nous avons développés pour la correction des QCM, ne permettent pas un gain de temps substantiel. Ils sont même relativement contraignants. Actuellement, hormis quelques enseignants prêts à s'investir dans la qualité docimologique, peu les utilisent. Notre prochain défi est sans doute d'améliorer leur utilisabilité et de mettre en place des accompagnements performants à cet égard.

Comme l'écrit Scallon (2004) L'évaluation des apprentissages dans une approche par compétences ressemble à un chemin sinueux, un chemin qui croiserait plusieurs sentiers propices à la distraction et parfois à la confusion.

Nous devons être attentifs à ne pas perdre trop de temps sur ce chemin !

Références

Argyris, C, Schön, A. D. (1978) *Organizational Learning: A Theory of Action Perspective*, Addison-Wesley in Reading, MA, "L'apprentissage organisationnel, théorie, méthode et pratique"

- Charlier E. (2000). *Développer la réflexivité: entre le dire et le faire*, in G. Carlier, J.-P. Renard, L. Paquay, Enjeux, innovation et réflexivité, Bruxelles : De Boeck, p. 111-120
- Charlin et al. (2005) *Le test de concordance comme outil d'évaluation en ligne du raisonnement des professionnels en situation d'incertitude*, Revue internationale des technologies en pédagogie universitaire, 2(1)
- De Ketele, J.-M. & Gérard, F.-M. (2005). La validation des épreuves d'évaluation selon l'approche par les compétences, *Mesure et Évaluation en Éducation*, Volume 28, n°3, 1- 26.
- Detroz, P., Crahay, V., Piazza, A. (2013) *Traitement des Questions à Réponse Ouverte Longue (TQROL) - Vers. 1.1 [Computer Software]*. Liège : Université de Liège, Institut de formation et de recherche en enseignement supérieur (IFRES), Unité de soutien logistique et de recherche « Système méthodologique d'aide à la réalisation de tests (SMART) ».
- Gilles, J.-L. (2002). *Qualité spectrale des tests standardisés universitaires – Mise au point d'indices éducatifs d'analyse de la qualité spectrale des évaluations des acquis des étudiants universitaires et application aux épreuves MOHICAN check up '99*, thèse de doctorat en Sciences de l'éducation. Liège : Université de Liège, Faculté de psychologie et des sciences de l'éducation de l'université de Liège (559 pages).
- Gilles, J.-L., Detroz, P., Crahay, V., Humblet, J.-P., Pirson, M., Lescalier, Y., Lê, H., Bruyère, F. (2008) *Examination Assessment Management System (ExAMS) - Vers. 1.1 [Computer Software]*. Liège : Université de Liège, Institut de formation et de recherche en enseignement supérieur (IFRES), Unité de soutien logistique et de recherche « Système méthodologique d'aide à la réalisation de tests (SMART) ».
- Gilles, J.-L., Detroz, P., Crahay, V., Tinnirello, S. et Bonnet, P. (2011). La plateforme ExAMS, un "assessment management system" pour instrumenter la construction et la gestion qualité des évaluations des apprentissages. Dans Blais, J.-G. et Gilles, J.-L. (Dir.), *Évaluation des apprentissages et technologies de l'information et de la communication – Tome II*. Québec : Presses de l'Université de Laval.
- Harden R, Stevenson M, Downie W, Wilson G. *Assessment of clinical competence using objective structured examination*. BMJ 1975 ; 1 : 447-451
- Kluger, A. N., & 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), 254–284
- Kulhavy, R. W., & Stock, W. (1989). Feedback in written instruction: The place of response certainty. *Educational Psychology Review*, 1(4), 279–308.
- Lambert, M., Rossier, A., Daele, A. & Lenzo, G. (2009). *Le feedback aux étudiant-e-s*. Université de Lausanne. Retrieved September, 30 from http://pedagogieuniversitaire.files.wordpress.com/2009/08/memento_feedback_29juillet2009_vl.pdf

- Leclercq, D. (1986). *La conception des QCM*. Bruxelles : Labor.
- Leclercq, D. (2006) L'évolution des QCM. In G. Figari et L. Mottier-Lopez. *Recherches sur l'évaluation en Education*. Paris : L'Harmattan, 139-146.
- Linn, R. L., & Grolund, N. E. (2000). *Measurement and assessment in teaching* (8th ed.). Upper Saddle River, NJ: Merrill/Prentice Hall.
- Nicol, D. & MacFarlane-Dick, D. (2007). Formative assessment and self-regulated learning : A model and seven principles of good feedback practice. *Studies in Higher Education*, 3(2), 199-218.
- Nitko, A.J. (2004). *Educational assessment of children* (4th ed.). Upper Saddle River, NJ: Pearson.
- Perrenoud (2001) *Mettre la pratique réflexive au centre du projet de formation* In Cahiers Pédagogiques, n° 390, pp. 42-45.
- Phelps, R. P., Ed. (2005a). *Defending standardized testing*. Mahwah, NJ, USA: Lawrence Erlbaum.
- Scallon, G. (2004). *L'évaluation des apprentissages dans une approche par les compétences*. Bruxelles : De Boeck Université.
- Schön D. (1994), *Le praticien réflexif. A la recherche du savoir caché dans l 'agir professionnel*, Logiques
- Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, 78(1), 153-189.
- Tardif, J. (2006). *L'évaluation des compétences. Documenter le parcours de développement*. Montréal: Chenelière.