

Application du cycle gestion qualité S.M.A.R.T. des tests pédagogiques au cours d'Obstétrique et de Pathologie de la Reproduction des ruminants, équidés et porcins.

Jean-Loup Castaigne¹, Jean-Luc Gilles², Christian Hanzen¹

Résumé : Suite aux résultats de l'évaluation de l'enseignement (EvalEns : Leclercq & Gilles, 1995 ; Gilles & Pirson, 1999 ; Boxus, 2000) réalisée auprès des étudiants de 2^{ème} doctorat durant l'année académique 1999-2000, nous avons décidé de mettre en place de nouvelles stratégies de formation et d'évaluation. La méthode des LQRT³-SAFE⁴ (Leclercq *et al.*, 1993 et 1998) a été choisie pour dispenser 5 cours portant sur des matières dont la compréhension est capitale pour la réussite de l'examen du professeur Hanzen. Chaque test a été réalisé selon les principes définis par le Système Méthodologique d'Aide à la Réalisation de Tests (SMART) dans leur cycle de gestion qualité des évaluations standardisées. Chaque test est découpé en 8 étapes clefs telles que définies par Gilles & Leclercq (1995). Des boucles de régulation courtes et longues assurent un maintien de la qualité par rapport aux objectifs d'évaluation définis par l'enseignant.

Introduction

La plupart du temps, les enseignants évaluent des processus mentaux appliqués à des contenus, or, s'il est aisé de mettre en évidence les contenus enseignés, il n'en va pas toujours de même en ce qui concerne les processus mentaux. Il existe cependant des outils méthodologiques qui permettent de clarifier ces processus mentaux.

Nous souhaitons évaluer la compréhension et le jugement critique de nos étudiants concernant 5 points de matière fixés par le professeur Hanzen. Pour clarifier les objectifs cognitifs, nous nous sommes référés à la taxonomie de Bloom (1969) qui distingue : (1) la connaissance c'est-à-dire la restitution de mémoire, (2) la compréhension ou interprétation correcte de données, de concepts et de raisonnements, (3) l'application de principes à la solution de cas classiques, (4) l'analyse ou la détection de problèmes et les classifications, (5) la synthèse, c'est-à-dire l'expression et la (re)formulation, (6) l'évaluation ou le jugement sur base de critères personnels. Selon Gilles et Leclercq (1995), dans les évaluations habituellement pratiquées, les niveaux (1) connaissance, (3) application et (5) synthèse sont mesurés plus systématiquement que les trois autres, c'est-à-dire de façon consciente, organisée et de manière telle que tous les étudiants se voient poser un échantillonnage de questions relevant de ces niveaux.

¹ Service d'Obstétrique et Pathologie de la Reproduction des ruminants, équidés et porcs et pathologie de la mamelle de la vache laitière, *Faculté de Médecine Vétérinaire, Université de Liège — Belgique*, Tél. : +32 – 4 – 366 41 60, Fax +32 – 4 – 366 41 36, eMail: oga.fmv@ulg.ac.be, Bât. B42 - OGA, boulevard de Colonster, 20, 4000 Liège Sart-Tilman; site web : <http://www.fmv.ulg.ac.be/oga/>

² Technologies de l'éducation, *Faculté de Psychologie et des Sciences de l'Education - Université de Liège - Belgique*; Tél. : + 32 - 4 - 366.20.78 - Fax : + 32 - 4 - 366.29.53, eMail : JL.Gilles@ulg.ac.be, Bât. B32, boulevard du Rectorat, 5, 4000 Liège 1

³ LQRT : Lire, Questions-Réponses, Test (Leclercq, 1998)

⁴ SAFE : Système Adulte de Formation et d'Évaluation (Leclercq, 1971)

Pour mesurer les trois autres niveaux de la taxonomie de Bloom que sont (2) la compréhension, (4) l'analyse et (6) l'évaluation, nous utilisons plusieurs procédures combinées comme recommandé par Gilles & Leclercq (1995). Ces procédures sont :

- les Questions à Choix Multiple (QCM) avec Solutions Générales Implicites (SGI) qui autorisent, en plus des solutions habituellement proposées, les quatre possibilités suivantes : Rejet (aucune solution proposée n'est correcte), Toutes (toutes sont correctes), Manque (il manque des données dans l'énoncé pour que l'on puisse choisir UNE solution comme correcte), Absurdité (il y a une contre-vérité dans l'énoncé à dénoncer en priorité !); (Wood, 1977 et Leclercq, 1993a,)
- les évaluations à livres ouverts (LO) (Leclercq *et al.*, 1993) et
- les degrés de certitude (DC) en respectant les 4 consignes⁵ définies par Shufford (1996) et Leclercq (1983, 1986, 1993b), que nous avons intégrés au modèle global de gestion des examens défini dans le cadre du SMART (Gilles & Leclercq, 1995). Le but de ce système est d'augmenter la validité et la fidélité des examens. Nous en proposons ci-dessous une version simplifiée que nous allons détailler point par point.

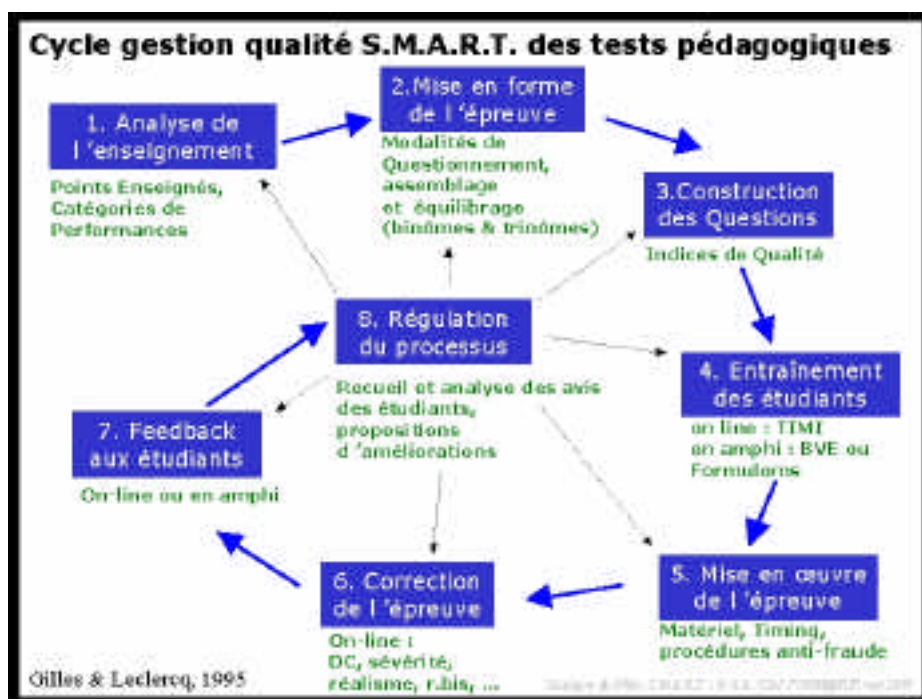


Figure 1: Cycle de gestion qualité SMART des tests pédagogiques

1. Analyse de l'enseignement

Le contenu du cours est découpé en modules ou points d'enseignement jugés suffisamment importants que pour en justifier d'une évaluation. Pour chaque point d'enseignement (PE), l'enseignant quantifie son importance relative et les catégories de performances (CP) attendues, autrement dit les mesures que l'enseignant désire lors de l'évaluation de l'apprenant. Comme CP, nous avons utilisé la taxonomie des objectifs cognitifs

⁵ (1) une consigne probabiliste, (2) un barème des tarifs calculé selon la théorie des décisions, (3) le calcul d'indice de réalisme, (4) un entraînement à la procédure.

de Bloom. Nous avons construit un premier binôme en associant à un point d'enseignement une catégorie de performance qui permettra d'en apprécier la bonne transmission vers l'apprenant. Chaque PE peut être associé à plusieurs CP en créant autant de binômes différents. Notre liste de binômes est constituée de [PE x CP] où les points d'enseignements sont des parties des notes et les CP les objectifs cognitifs tels que définis par Bloom. La liste des points enseignés et leur importance relative est communiquée aux étudiants au travers d'une page Web dédiée (<http://www.fmv.ulg.ac.be/oga/aide/lqrt2/PE2.html>).

2. Mise en forme de l'épreuve

Pour cette partie, nous utilisons le kaléidoscope des techniques de questionnement tel que décrit par Gilles & Leclercq (Leclercq, 1999, chapitre VI). Nous avons choisi comme modalité de questionnement les QCM SGI en y associant 2 conditions : « examen à livres ouverts » et « avec degré de certitude ».

Les QCM ont été choisies pour réaliser une évaluation standardisée et pour la facilité de correction des grands groupes (150 étudiants). L'évaluation sera complétée par un examen oral en fin d'année portant sur l'ensemble de la matière.

Les SGI nous permettent d'évaluer des niveaux taxonomiques plus élevés que la connaissance dans la taxonomie de Bloom, d'éviter l'identification de la réponse correcte (recognition) par hasard, de mesurer la détection d'erreurs, la vigilance factuelle⁶, de combattre le curriculum caché⁷ et d'habituer aux situations où plusieurs solutions sont correctes.

L'évaluation se déroule à livre ouvert car il nous paraît évident qu'une épreuve ne portant pas sur l'évocation et le rappel devrait se dérouler à livre ouvert.

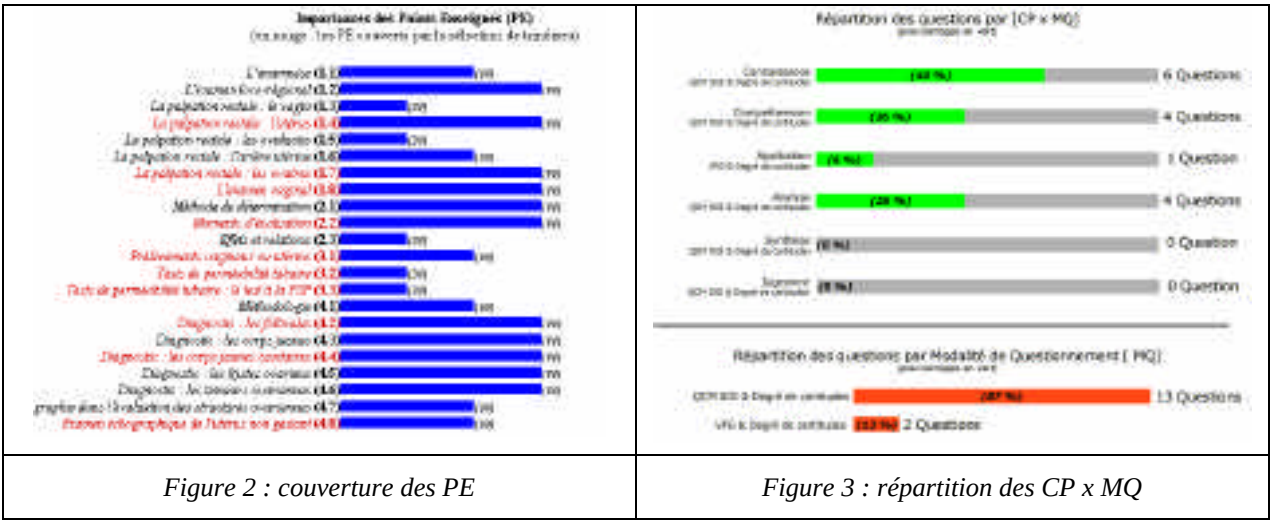
Pour mesurer la métacognition, demander à l'étudiant d'exprimer sa certitude nous permet d'une part de l'amener à s'auto-estimer (niveau 6 dans la taxonomie de Bloom). D'autres part en comparant sa prédiction et son résultat, nous calculons ses indices de réalisme, de centration et de cohérence.

Nous avons ainsi défini une modalité de questionnement (MQ) : « mesurer la compréhension à l'aide de QCM SGI avec degrés de certitude et à livres ouverts ». Pour chaque catégorie de performance, nous pouvons choisir une modalité de questionnement et créer de la sorte une seconde liste de binômes [MQ x CP]. Nous pouvons croiser cette liste de binômes avec la première en utilisant comme dénominateur commun les catégories de performance (CP) et obtenir une liste de trinômes [PE x CP x MQ] qui nous permet déjà de vérifier la cohérence de notre épreuve en étudiant la répartition des questions en fonction des PE et des CP. Le résultat de

⁶ Aller spontanément au-delà du donné, accéder à l'implicite, détecter les pièges, les incohérences, les lacunes, les contradictions sans être mis sur la voie et donc lutter contre les messages implicites tel que « Quand on vous pose une question, il faut répondre », « Quand une question est posée, il existe une et une seule bonne réponse » ou « Une question posée par l'autorité est forcément bien posée ». (Leclercq, 1999)

⁷ Ce que personne n'enseigne, mais que tout le monde apprend (Leclercq, 1999)

ces choix et opérations est synthétisé dans les figures 2 et 3.



3. Construction des questions

Sur base des trinômes, nous pouvons générer des questions correspondant aux objectifs que nous nous sommes fixés. Pour la rédaction des QCM-SGI, nous nous référons aux principes définis par Leclercq (1986). Le professeur et les assistants rédigent chacun une série de questions qui sont présentées à l'ensemble des encadrants du service qui essaye d'y répondre. S'en suit une discussion dans laquelle l'auteur défend sa question, la modifie ou la supprime le cas échéant. Une vérification du respect des règles de conception des questions est faite en même temps. Une sélection des trente meilleures questions dont 50 % ont comme solution une SGI (Leclercq, 1986 et MOHICAN) est effectuée et proposée à chaque membre du staff pour une relecture formelle.

Les étudiants de deuxième doctorat de notre Faculté n'ont rencontré qu'une seule fois durant leur cursus les examens sous forme de QCM dans lesquels seules les SGI "aucune" et "toutes simultanément" étaient présentes et ceci sans avoir reçu aucun entraînement préalable. Suivant les recommandations de Leclercq et collaborateurs (1993), les autres SGI ont été introduites progressivement. Sur 30 QCM, 50 % avaient comme réponse correcte une SGI. Sur ces 15 QCM, le premier test comportait deux SGI "absurdité" et aucune SGI "manque de données". Le deuxième test contenait 3 SGI "manque de données" et 4 SGI "absurdité". Leclercq a montré en 1993 que la SGI "manque de données" était la plus complexe à appréhender pour les évalués en Faculté de Psychologie et Sciences de l'Éducation et c'est pourquoi nous avons d'abord ajouté la SGI 9 avant la SGI 8. Ce n'est pas le cas pour toutes les matières. Ainsi, Jans (1994) a montré que les enfants de 9 à 10 ans qui rédigent eux-mêmes des QCM-SGI en mathématique, construisaient des QCM-SGI "manque" faciles à détecter par leurs condisciples.

Les analyses des résultats nous ont montré qu'à partir du deuxième test, les étudiants réussissaient aussi brillamment les questions dont la solution était une des propositions que celles dont la solution était une SGI. Nous nous sommes dès lors concentrés plus sur la qualité des SGI que sur leur quantité respective, choisissant une SGI en fonction du processus mental que nous voulions que l'étudiant réalise en répondant.

4. Entraînement des étudiants

Une grande importance a été donnée aux entraînements. Chaque semaine un groupe d'une douzaine d'étudiants recevait une séance d'information de trois heures sur ordinateur afin de les familiariser avec les QCM-SGI-DC. Spécifiquement pour les degrés de certitude, les étudiants ont reçu une information théorique expliquant notamment le réalisme et la centration avant de passer à une mise en pratique individuelle grâce au logiciel GUESS (Leclercq et Gilles, 1994) où les étudiants sont invités à deviner les lettres successives d'un texte tiré au hasard dans la littérature française, chacune de leur réponse étant accompagné d'un degré de certitude. L'étudiant voit se constituer immédiatement son graphe de réalisme (voir figure 4). L'exercice était axé sur l'utilisation de toutes les classes de certitudes, le réalisme et la centration.

Une partie de la formation hebdomadaire était consacrée à l'explication des objectifs des SGI et leur mise en pratique au travers des 35 QCM-SGI-DC-LO portant sur la matière des candidatures et le programme des cours de notre Faculté. L'exercice portait toujours sur l'expression de la certitude mais en y ajoutant les SGI puisque la réponse correcte à 9 questions sur 10 était une SGI. Les étudiants recevaient un feedback immédiat grâce au logiciel TIMI⁸. (figure 5).

Les feedbacks fournis aux étudiants par le logiciel TIMI portent sur deux points principaux :

- Leur réalisme et centration, la répartition des réponses dans chaque classe de certitude
- La réponse correcte à la QCM avec une courte explication

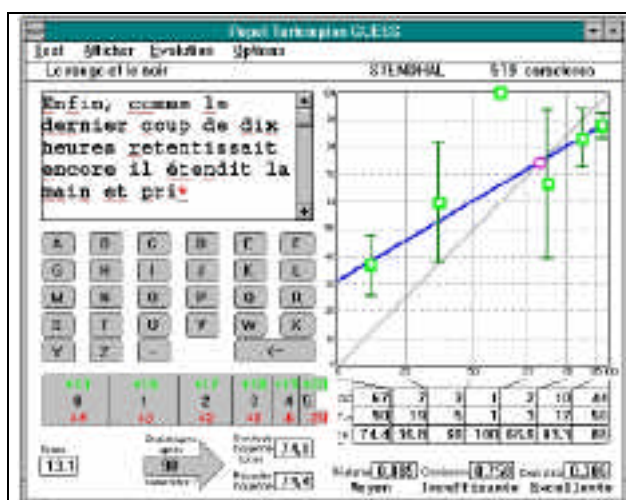


Figure 4 : GUESS

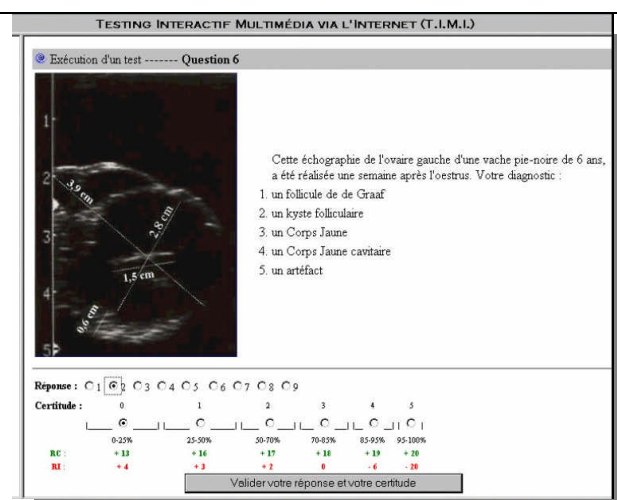


Figure 5 : TIMI

Avant chaque LQRT-SAFE, 6 QCM-SGI-DC-LO spécifiques à la matière étaient mise *on-line*. En respectant l'ordre de cet entraînement nous avons ainsi préparé progressivement les étudiants à aborder les questions propres à la matière en éveillant leur vigilance cognitive et leur métacognition.

⁸ TIMI : Testing Interactif Multimédia via l'Internet, Jean-Luc Gilles et M'Bassa Dabo, site web : <http://www.smart.ulg.ac.be/>

5. Mise en œuvre de l'épreuve

Les tests n'étant qu'une façon de forcer les étudiants à avoir lu de manière attentive la matière pour les questions réponses, le LQRT devait se dérouler sur une tranche de deux heures d'affilée : une heure de questions-réponses et une heure d'évaluation certificative à l'aide de 30 QCM-SGI-DC-LO. Il a fallu échanger des heures de cours pour mettre en place cette nouvelle méthode d'enseignement pour notre Faculté. Aucun amphithéâtre de notre faculté n'ayant la capacité suffisante pour pouvoir organiser une évaluation certificative pour les 150 étudiants que compte le deuxième doctorat, il a fallu déménager pour occuper un plus grand amphithéâtre au Centre Hospitalier Universitaire du Sart Tilman.

Le timing de chaque séance était divisé en deux parties : une heure de questions des étudiants auxquelles le professeur Hanzen répondait en illustrant ses propos de diapositives ou de séquences vidéo projetées sur grand écran et une heure pour le test. Une pause de dix minutes était instaurée entre les deux. Les assistants et le professeur répondaient individuellement aux questions des étudiants durant l'épreuve.

À la fin du test, les réponses correctes étaient affichées et les formateurs recueillaient les commentaires des étudiants.

Du point de vue matériel, les étudiants recevaient un questionnaire de 6 pages sur lesquelles étaient imprimées les définitions des SGI et le barème des tarifs (Leclercq, 1998) en fonction de chaque classe de certitude (voir illustration ci-dessous). Les étudiants étaient invités à répondre sur formuloms (formulaires pour la lecture optique de marques) en noircissant la case correspondant à leur réponse et à leur certitude. Quatre formes différentes du questionnaire étaient distribuées pour limiter la fraude.

SGI : 6=aucune 7=toutes 8=manque dans l'amorce 9=contre-vérité dans l'amorce

Votre certitude	0-25%	25-50%	50-70%	70-85%	85-95%	95-100%
cochez	0	1	2	3	4	5
réponse correcte	+ 13	+ 16	+ 17	+ 18	+ 19	+ 20
réponse incorrecte	+ 4	+ 3	+ 2	0	- 6	- 20

Figure 6 : bas de page des questionnaires remis aux étudiants

Chaque étudiant recevait en plus une feuille de justifications sur laquelle il était libre de justifier des réponses sachant que le correcteur ne lirait que les commentaires concernant les réponses incorrectes. La justification ne peut donc QUE bénéficier à l'étudiant.

6. Correction de l'épreuve

Une fois lues les feuilles destinées à la lecture optique, les réponses sont traitées par le logiciel CERT⁹. Le logiciel CERT fournit l'analyse de chaque question de l'épreuve en trois lignes par question.

Certm										
Processus : Pas de sélection.					Matière : Pas de sélection.					
	SOL: 0	1	2	3	4	5	6	7	8	9
Q 26	0.00	26.40	3.37	7.30	33.71	5.62	11.24	1.69	1.12	9.55
Rbis	0.00	0.01	0.11	-0.27	0.25	-0.03	-0.03	-0.04	-0.12	-0.14
Cmoy	0.00	41.44	32.50	42.31	49.21	33.50	37.63	29.17	45.00	49.71
Q 27	0.56	17.42	1.69	24.16	8.99	10.67	18.54	14.04	2.81	1.12
Rbis	-0.18	0.07	0.00	0.09	-0.05	-0.13	0.04	0.00	-0.09	0.05
Cmoy	12.50	37.26	37.50	38.95	36.41	31.84	44.07	44.30	36.50	12.50
Q 28	0.00	1.12	2.81	44.94	26.97	10.67	8.77	0.00	1.69	2.81
Rbis	0.00	-0.08	-0.10	0.25	-0.14	-0.14	0.06	0.00	-0.03	-0.04
Cmoy	0.00	25.00	12.50	59.03	53.13	40.00	46.41	0.00	46.67	32.00
Q 29	0.00	3.93	2.25	3.93	8.99	68.54	3.93	0.00	0.56	7.87
Rbis	0.00	-0.24	-0.09	-0.15	-0.11	0.39	-0.13	0.00	-0.09	-0.12
Cmoy	0.00	42.50	56.25	40.36	47.97	61.37	51.79	0.00	60.00	56.96

Figure 7 : fenêtre de présentation des résultats de CERT

Pour chaque solution proposée sont fournis : (1) le pourcentage d'étudiants qui ont choisi la proposition (ligne Q ii), (2) le coefficient r.bis (Rbis) et (3) la certitude moyenne (Cmoy). Le coefficient de corrélation bisériale de point (r.bis) est utilisé comme indice de corrélation entre le choix des propositions d'une QCM et les résultats globaux de l'épreuve. Lorsqu'une QCM « fonctionne » bien, on s'attend à un r.bis positif pour la réponse correcte et des r.bis négatifs pour les autres propositions. Le seuil à partir duquel le r.bis d'une réponse correcte peut être considéré comme satisfaisant est fonction du nombre de questions de l'épreuve (l'inverse de la racine carrée du nombre de questions (n) soit $\frac{1}{\sqrt{n}}$). Donc moins il y a de questions, plus le seuil est élevé.

L'analyse d'une épreuve à l'aide de ce dispositif nous a amené à supprimer certaines questions, par exemple lorsque la réponse correcte récolte un r.bis négatif élevé et les distracteurs des r.bis positifs. Nous avons aussi occasionnellement décidé d'accepter le choix d'un distracteur comme réponse correcte unique ou supplémentaire au vu des r.bis et / ou des justifications des étudiants. La question a alors été remaniée avant d'être réintroduite dans la banque de questions. Les justifications revenant le plus fréquemment étaient discutées lors d'une réunion de service à la fin de laquelle il était décidé si une réponse correcte était accordée à tous les étudiants ou uniquement aux étudiants ayant justifié correctement leur choix. Le degré de sévérité de l'épreuve était alors choisi en fonction de la difficulté apparente exprimée oralement par les étudiants, de la distribution des résultats, de l'indice de réalisme et l'indice de centration (Leclercq, 1993b).

⁹ CERT : logiciel réalisé en 1991 par Jean-Luc Gilles, sous les auspices de la Commission des Communautés européennes dans le cadre du programme EuroTecNet (Boxus *et al.*, 1991)

Grâce à la rapidité de la lecture optique des formuloms, des scores (basés sur d'autres critères que la réponse correcte et la certitude) comme par exemple la partie de matière ou le processus mental, ont pu être calculés. Pour les processus mentaux, l'évaluation portant principalement sur la compréhension, nous avons utilisé ce paramètre pour mesurer les résultats aux QCM dont la réponse correcte était une proposition dactylographiée (de 1 à 5) par opposition aux QCM dont la réponse correcte était une SGI (figures 8 et 9). Nous avons pu fournir à chaque étudiant un feedback supplémentaire pour lui permettre d'adapter ses stratégies en vue de la prochaine évaluation.

En ce qui concerne les matières enseignées, nous les avons regroupé les points enseignés en 3 ou 4 groupes de façon à pouvoir isoler les faiblesses et les forces de chacun dans telle ou telle partie des notes. Nous les avons également utilisés pour mesurer la distribution des cotes en fonction de l'auteur des questions (figures 10 et 11).

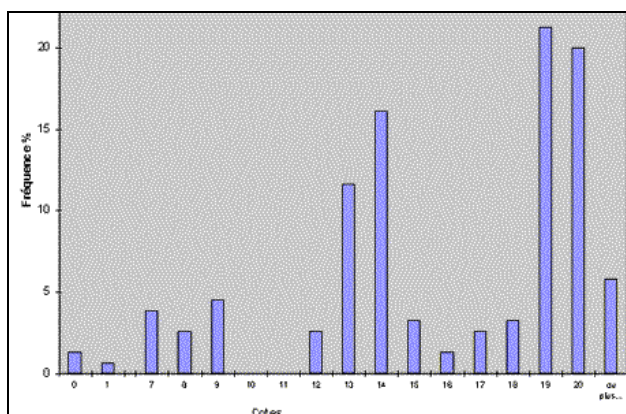


Figure 8 : Exemple de distribution des résultats en fonction du type de réponse correcte (non SGI)

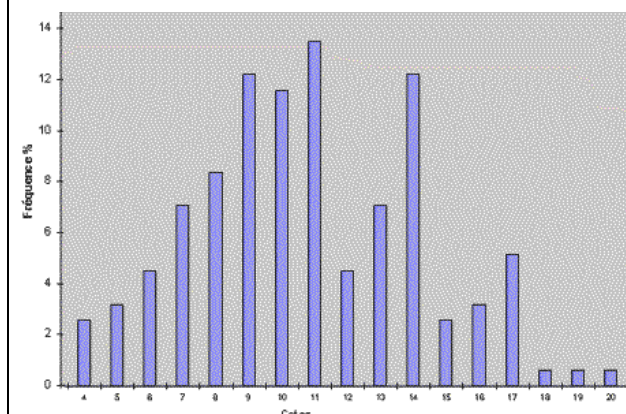
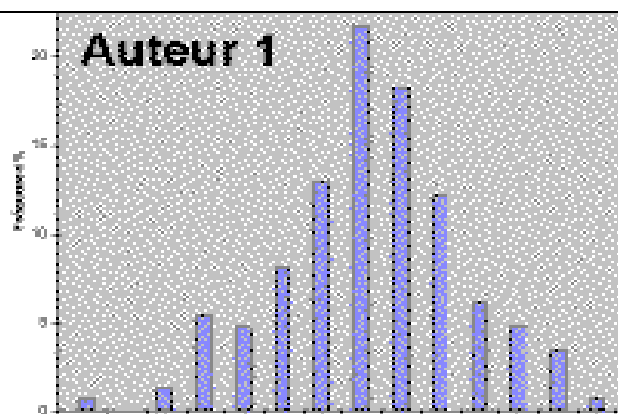
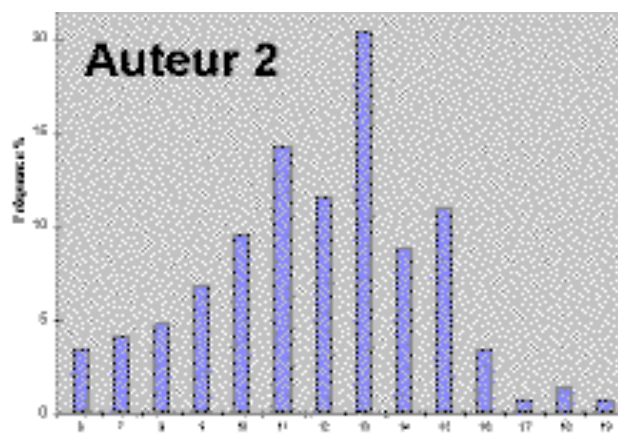


Figure 9 : Exemple de distribution des résultats en fonction du type de réponse correcte (SGI)



Figures 10 et 11 : Exemples de distribution des résultats en fonction de l'auteur ; note en ordonnée et fréquence d'apparition en abscisse)



Des traitements statistiques plus poussés seront réalisés quand les 5 LQRT et l'examen final auront eu lieu.

7. Feedbacks aux étudiants

a) Par le Web

L'intégralité du processus était décrite aux étudiants sur le site Web du service du professeur Hanzen. Les étudiants y trouvaient les résultats des analyses de r.bis, les commentaires concernant les justifications et les histogrammes de distribution des notes afin de pouvoir se situer par rapport au groupe. Chaque étudiant pouvait consulter sur le Web ses résultats qui comprenaient son score, son graphe de réalisme, ses indices de réalisme, de cohérence et de centration, la distribution de ses certitudes par classe, l'ensemble de ses réponses associées aux certitudes et les réponses correctes.

b) En présentiel

L'heure de cours suivant le LQRT était également consacrée à un feedback en amphithéâtre. Dans un premier temps, nous reprenions les QCM ayant posé des problèmes tels que des r.bis de distracteurs significativement positifs ou un nombre important d'étudiants ayant choisi une mauvaise réponse. Les étudiants discutant plus pour faire admettre chacun leur avis personnel que pour comprendre la matière qui était sous-jacente, nous avons changé de paradigme pour consacrer cette heure de cours à la ré-explication des points enseignés pour lesquels une majorité d'étudiants avaient échoué lors du test par le professeur Hanzen sous forme de réception-transmission (Leclercq et Denis, 1999).

8. Régulation du processus

Le processus n'attend pas d'être arrivé à de cette étape de régulation pour s'autoréguler. Plusieurs régulations ont déjà eu lieu, une étape pouvant se réguler elle-même ou vérifier la qualité d'une étape précédente. Dans notre cas par exemple, nous avons observé les régulations suivantes :

- a) À la mise en forme de l'épreuve (étape 2), nous avons validé les questions par rapport aux objectifs (étape 1).
- b) Dès la discussion en fin de test (5), nous avons changé les réponses correctes de certaines QCM (3).
- c) À la fin du premier LQRT (5), nous avons autorisé les étudiants à bénéficier d'un quart d'heure supplémentaire pour répondre au test (5).
- d) Lors de la correction des formuloms et des justifications (6), des QCM ont été supprimées du test et modifiées ou supprimées dans la banque des questions (3), d'autres solutions ont été admises comme correctes (6) et les fréquences de distribution des SGI ont été adaptées en fonction de la maîtrise montrée par les évalués (2).
- e) Les étudiants utilisant TIMI sans lire les feedbacks mais uniquement en espérant qu'une ou plusieurs QCM d'entraînement (4) seraient également dans le test, la forme et la réponse correcte des questions d'entraînement étaient différentes des questions du test (3).
- f) Une nouvelle EvalEns sera organisée après l'examen oral de fin d'année pour recueillir l'avis des étudiants et essayer d'améliorer le système pour l'année académique suivante.

La régulation a donc bien lieu à tous les niveaux, certains niveaux s'autorégulant, d'autres validant les objectifs des points précédents. Un niveau de régulation du processus complet consiste à regrouper les régulations ayant déjà eu lieu et à y ajouter l'analyse des avis des évalués

afin de proposer des améliorations remettant en cause l'exécution de la méthode (boucles de régulations courtes portant sur les points 4 à 7 de notre schéma) mais aussi la conception même de la méthode (boucles longues portant sur les points 1 à 3 mais aussi sur l'intérêt de poursuivre ou non dans cette voie).

Conclusions

L'équipe du SMART travaille à automatiser par l'informatique le cycle gestion de qualité des tests pédagogiques. Un certain nombre d'étapes sont à l'heure actuelle gérables *on-line* par l'enseignant. De nouveaux outils sont encore à créer et à intégrer. Ils devraient aider les enseignants à clarifier non seulement les contenus à évaluer, mais aussi les processus mentaux et par là leurs objectifs. Ils devraient également contribuer à faire le lien avec la forme d'examen la plus appropriée (Gilles & Leclercq, 1995). L'adjonction d'outils d'évaluation de la qualité des questions comme le r.bis spectral et r.bis spectral contrasté turbo participerait à l'amélioration de la fiabilité des questions. L'analyse spectrale est basée sur la sélection des étudiants les plus réalistes c'est-à-dire ceux qui obtiennent les meilleurs scores à l'indice de réalisme. Ensuite on calcule les r.bis de toutes les propositions uniquement sur base de cette sélection d'étudiants (Gilles, 1999). Enfin l'analyse spectrale des résultats et leur présentation sous forme d'hémispectres des réponses correctes et incorrectes (Jans et Leclercq, 1999) permettrait de communiquer aux étudiants des informations concernant leurs connaissances utilisables, inutilisables ou dangereuses.

Les aspects liés à l'interdisciplinarité ne sont pas à négliger : beaucoup de contenus de cours différents se rejoignent et c'est heureux. Des pratiques de partage des cours devraient amener plus de cohérence dans les cursus et aider les étudiants à mieux percevoir les relations qu'entretiennent les différentes disciplines relevant d'un même domaine. La partie formation de l'apprentissage pourrait être réorganisée lors du processus de départementalisation des services qui est en cours à l'ULg. Cela implique un regroupement des services en fonction des contenus relevant d'un même domaine. Dès lors nous pourrions instaurer un système alternatif dans lequel certains professeurs seraient responsables une année de la partie formation et d'autres professeurs responsables de la partie évaluation et d'inverser les rôles l'année suivante.

Pour améliorer la validité, Gilles conseille de demander plus fréquemment aux étudiants leur avis sur la représentativité des questions. Comment perçoivent-ils les liens entre les questions posées à l'examen et les objectifs de l'enseignement qu'ils ont suivi ? Quels sont leurs commentaires à ce sujet ? Récoltés systématiquement à l'occasion de chaque examen et pris en compte par les enseignants, ces avis amélioreraient la validité des épreuves.

De façon plus globale, en vue d'améliorer les cours, il est intéressant de récolter les avis à propos du processus d'enseignement et d'établir les liens entre les résultats, les caractéristiques des cours, les objectifs, les modalités d'évaluation et la perception qu'ont les étudiants et les enseignants de l'ensemble.

Enfin, le problème de la transférabilité des compétences acquises au cours des études et certifiées lors des évaluations doit, à terme, également être pris en compte. L'évaluation de la qualité de la formation devrait pouvoir se poursuivre auprès des diplômés. Une étude de ce type est en cours pour les vétérinaires installés en Belgique. En ne nous donnant pas les moyens

d'estimer la qualité des transferts nous nous privons d'informations primordiales pour un enseignement cohérent et connecté sur la réalité professionnelle.

-

Bibliographie

- BLOOM, B. *et al.* (1969). Taxonomie des objectifs pédagogiques. I. Domaine cognitif, traduit par M. Lavallée, Montréal, Education Nouvelle.
- BOXUS *et al.* (1991). Principes communs pour évaluer les résultats cognitifs de la formation. Commissions des Communautés européennes, programme Eurotecnet.
- BOXUS E. *et al.* (2000). AEE/CGE Commission Évaluation des Enseignements - Synthèse Générale, Université de Liège : mai 2000.
- GILLES J.-L. (1999). Apports des mesures métacognitives lors d'un test de sélection portant sur la compréhension d'un article scientifique en 1^{ère} candidature de la faculté de médecine, in Ch. Depover et B. Noels (eds), *Approches plurielles de l'évaluation des compétences des processus cognitifs*, Proceedings of the 12th ADMEE Conference, Mons ; UMH-FUCAM, 19-30
- GILLES J.-L. & LECLERCQ, D. (1995). Procédures d'évaluation adaptées à de grands groupes d'étudiants universitaires : enjeux et solutions pratiquées à la FAPSE-ULG, Actes du symposium International sur la rénovation didactique en biologie – novembre 1995 – Université de Tunis
- GILLES J.-L. & PIRSON M. (1999). Document de synthèse pour la Commission Inter-Facultaire d'Évaluation des Enseignements (campagne 1997-1998) ; Université de Liège : Évaluation des Enseignements, Données publiques
- JANS, V (1994). Développer l'autoestimation et la vigilance cognitive à l'école primaire, mémoire de licence en Sciences et Techniques de la formation continue, FAPSE, ULg
- JANS, V. & LECLERCQ, D. (1999). Mesurer l'effet de l'apprentissage à l'aide de l'analyse spectrale des performances, in Ch. Depover et B. Noels (eds), *L'évaluation des compétences et des processus cognitifs. Modèles, pratiques et contextes*, Bruxelles :DeBoeck , 1999, 303-317
- LECLERCQ, D. (1983). Confidence marking, its use in testing. Postlethwaite, Choppin (eds.) *Evaluation in Education*, Oxford : Pergamon, 1982, vol. 6, 2, pp. 161-287.
- LECLERCQ, D. (1986). La conception des questions à choix multiple, Bruxelles, Ed. Labor.
- LECLERCQ, D. *et al.* (1993a). The Taste approach: General implicit solutions in MCQs, open books exams and interactive testing and self-assessment. NATO ASI Series, Item Banking: Interactive Testing and Self Assessment, Berlin: Springer Verlag, 1993, Vol. 112, pp. 210-232.
- LECLERCQ, D. *et al.* (1993b). Validity, Reliability, and Acuity of Self-Assessment in Educational Testing. NATO ASI Series, Item Banking: Interactive Testing and Self Assessment, Berlin: Springer Verlag, 1993, Vol. 112, pp. 114-131.
- LECLERCQ, D (1999). Édu-métrie et docimologie, Liège : STE-ULG, 1999
- LECLERCQ, D *et al.* (1998). L'amphithéâtre électronique – Une application : le LQRT-SAFE in (Leclercq ed) *Pour une pédagogie universitaire de qualité*, Liège : Mardaga, 1998
- LECLERCQ, D et DENIS, B (1999). Méthode de formation et psychologie de l'apprentissage, Liège : STE-ULG, 1999
- LECLERCQ, D. & GILLES J.-L. (1994). GUESS, un logiciel pour entraîner à l'auto-estimation de sa compétence cognitive. Actes du colloque QCM et questionnaires fermés, Paris: ESIEE, 1994.
- LECLERCQ, D. & GILLES J.-L. (1995). EVALENS, Propositions pour un logiciel d'évaluation des enseignements. Commission d'évaluation des enseignements de la Faculté de Psychologie et des Sciences de l'Education, Université de Liège.
- MOHICAN, Monitoring Historique des étudiants de Candidature, Bruxelles : Conseil inter universitaire francophone, site CIUF : <http://www.ciuf.be/publications/reussitedoc.htm>
- SHUFORD, E. *et al.* (1966). Admissible Probability Measurement Procedures, *Psychometrika*, 31, 125-145
- WOOD R. (1977). Multiple choice : A state of the art report. In Chopin & Postlewaite (eds), *Evaluation in Education International Progress*. Oxford : Pergamon