# Impact du recours à une table de spécification sur la construction de test à choix multiples appliqué à la pathologie de la reproduction en médecine vétérinaire

# AIPU2008-05-0322

Ch. Hanzen (1), L.Théron (1), J.Sterkendries (1), V.Crahay (2)
Université de Liège
(1) Faculté de Médecine Vétérinaire
Service d'Obstétrique et de Pathologie de la Reproduction
B42 Sart Tilman, B-4000 Liège
Courriel : christian.hanzen@ulg.ac.be
Site Web : http://www.fmv.ulg.ac.be/oga/index.html
(2) Système Méthologique d'Aide à la Réalisation de Tests, Université de Liège
B3c Sart Tilman, B-4000 Liège
Site Web : http://www.exams.be/

## 1. Introduction

La faculté de médecine vétérinaire de l'Université de Liège se trouve confrontée depuis plusieurs années à un important problème de pléthore d'étudiants. Il en a résulté un recours pratiquement généralisé à des questionnaires à choix multiples (QCM) comportant dans certains cas des solutions générales implicites (SGI)[1] pour l'évaluation des étudiants. Comme d'autres méthodes d'évaluation, ce procédé présente tout à la fois des avantages et des inconvénients. Il permet en effet d'évaluer plusieurs niveaux d'activité mentale relatifs à l'ensemble de la matière[2]. Il permet de supprimer les effets liés au correcteur. Le recours aux formulaires optiques de marque autorise une correction rapide. A contrario, la rédaction des questions implique un temps certain surtout si elles concernent un niveau d'activité mentale plus élevé. Le risque de questions de détail est réel. Des mesures spécifiques doivent être prises pour éviter les fraudes lors de l'évaluation. Des propositions erronées risquent d'engendrer leur mémorisation à moins qu'une rétroaction soit délivrée sitôt l'évaluation réalisée (Karraker, 1967). Leur rédaction suppose enfin le respect de règles de base (Leclercq, 1986) pour éviter d'orienter favorablement ou non l'apprenant.

Nous avons recours depuis plusieurs années aux QCM-SGI attendu le nombre élevé d'étudiants que nous devons évaluer. Souhaitant nous inscrire d'emblée dans une approche aussi qualitative que possible de ce type d'évaluations, nous avons intégré en 2001 le modèle global de gestion des examens défini en 1995 par le Service d'Aide Méthodologique à la Réalisation de Tests (SMART) de l'Université de Liège (Gilles et Leclercq, 1995) en vue d'augmenter la validité et la fidélité des examens (Castaigne *et al*. 2001). Ce modèle a évolué au cours des années (Gilles 2002, Gilles *et al*., 2005) vers un cycle de Construction et Gestion de Qualité des Tests Standardisés (CGQTS) (http://www.smart.ulg.ac.be/smartweb/missions/evaletu_scientifique.php).

Le recours systématique depuis plusieurs années aux QCM-SGI nous a incité à formuler une double réflexion.

La première concerne le fait de savoir si ce type de questions n'induisait pas une méthode d'apprentissage trop basique car trop orientée sur la seule connaissance et beaucoup moins sur la compréhension, l'application voire l'analyse (Bloom 1968). A défaut d'enquête circonstanciée auprès des étudiants, on peut néanmoins raisonnablement le supposer. Les conséquences négatives de cette stratégie sont réelles. En effet, l'investissement dans la matière se faisant à court terme, la mémorisation à moyen voire long terme fait défaut. De plus, l'apprentissage est davantage vertical qu'horizontal ce qui ne permet pas à l'apprenant de faire des liens entre les divers aspects d'un même cours ni entre les cours. Par ce constat, nous supposons donc qu'il ne peut qu'en résulter une diminution des compétences. La solution apportée mais qui cependant ne peut être considérée comme définitive a été de rédiger des objectifs spécifiques. Ces objectifs ont été revus au fur et à mesure des années. C'est ainsi que lors de l'année académique 2007-2008, 200 objectifs ont été énoncés pour les 12 heures de cours de propédeutiques spéciales dispensées en 1er doctorat (4ème année d'un cursus qui en comporte six). De même, les 45 heures de cours organisées en 2ème doctorat ont fait l'objet de 300 objectifs. Afin de savoir si ces objectifs ont aidé les apprenants à assimiler la matière, nous leur avons demandées cette année encore, au travers d'un questionnaire de satisfaction, leur avis. Des 150 réponses reçues sur les 210 étudiants du cours, il ressort que 25 % n'ont pas exprimé d'opinion sur l'utilisation de ces objectifs pour leur apprentissage, que respectivement 49 % et 13 % étaient d'accord voire tout à fait d'accord avec la proposition et que respectivement 3 et 11 % n'étaient pas du tout ou pas d'accord avec la proposition. Ces pourcentages confirment les résultats observés les années précédentes. Cependant bien que répondant à un souci d'amélioration des apprentissages, la formulation de ce type d'objectifs pourrait avantageusement être remplacée par celle d'objectifs d'évaluation.

---

[1] Les Questions à Choix Multiple (QCM) avec Solutions Générales Implicites (SGI) autorisent, en plus des solutions habituellement proposées, les quatre possibilités suivantes qualifiées d'implicites car systématiquement proposées pour chaque question : 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, Leclercq 1993a)

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 factuelle [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)], de combattre le curriculum caché [ce que personne n'enseigne, mais que tout le monde apprend (Leclercq 1999)] et d'habituer aux situations où plusieurs solutions sont correctes.

[2] En effet, les QCM n'évaluent pas uniquement la connaissance. Nous pouvons également mesurer avec des QCM la compréhension, l'application et l'analyse de la taxonomie de Bloom.

La seconde réflexion concerne la nature des rétroactions délivrées aux apprenants. A la différence de bien d'autres collègues, nous avons toujours communiqué aux apprenants les réponses correctes et leurs justifications sitôt le test certificatif effectué. Cependant, cette façon de faire ne permet qu'imparfaitement d'identifier les causes d'un échec éventuel. Si une évaluation comprend divers niveaux d'activité cognitive tels la connaissance, la compréhension et l'application voire la métacognition, alors il semble normal qu'une rétroaction en fasse part également.
La mise au point par le SMART d'un système informatisé (Plateforme ExAMS – Exams Assessment Management System) de Construction et Gestion de Qualité des Tests Standardisés (CGQTS) nous a dans ce double contexte réellement interpellé. La présente communication se propose de faire état de l'approche de cet outil, des difficultés rencontrées et des enjeux auxquels il permet de faire face. Cette première expérience menée durant l'année académique 2007-2008, concerne 230 apprenants de 2ème doctorat en médecine vétérinaire.

## 2. Matériel et méthodes

### *2.1. La plateforme Exams : données générales*

La plateforme ExAMS conçue dans une philosophie « *Open Source* » soutient une démarche structurée en évaluation axée sur des concepts scientifiques innovants (décrite dans le point 2.3) apportant une plus value en termes de qualité à l'ensemble des acteurs.
Pour les évaluateurs, ExAMS offre une aide à la prise de décision sur l'ensemble de la démarche de création, gestion et correction de l'évaluation des apprentissages à l'aide d'une méthodologie en plusieurs étapes.
Pour les évalués, une communication optimale est assurée à travers la mise en place d'informations précisant le contrat docimologique, d'un monitoring précis en cours d'évaluation et de feedbacks diagnostiques personnalisés.
Pour les responsables institutionnels, l'usage de la plateforme garantit une meilleure qualité notamment du point de vue de la validité, fidélité, équité, ... des évaluations, qu'elles soient certificatives ou formatives.

### *2.2. L'enseignement concerné*

L'expérience concerne les 50 heures de cours données en présentiel dans le domaine de la pathologie obstétricale, de la reproduction et de la glande mammaire des ruminants, des équidés et des porcs. Les contenus Word et Power Point de ces matières sont mis à disposition des apprenants dès le début de l'année au moyen de la plateforme d'enseignement à distance WebCT utilisée pour la première fois en 2005-2006. Chacun des 36 chapitres du cours fait par ailleurs l'objet d'un forum de discussions et de tests formatifs. Selon les cas, l'un ou l'autre chapitre se trouve illustré par du multimédia.
Deux tests certificatifs sont organisés le premier en novembre et le second en janvier. Ces deux tests comportent 40 questions de type QCM-SGI et une voire deux questions dites à réponse ouverte longue (QROL). Le premier test concerne la matière vue depuis le début de l'année (30 % de l'ensemble), le second test concerne l'ensemble de la matière. La participation aux tests certificatifs et formatifs n'a aucun caractère obligatoire. Comme les années précédentes, l'apprenant pouvait, en cas de réussite aux deux tests certificatifs (moyenne égale ou supérieure à 12), reporter sa note pour l'examen oral de fin d'année et en être ainsi dispensé.

### *2.3. Le cycle de qualité*

Tel que défini, le cycle de Construction et de Gestion de la Qualité de Tests Standardisés (CGQTS) comporte 8 étapes qui sont l'analyse, le design, les questions, l'entraînement, le test proprement dit, sa correction, les feedbacks et la régulation.
La première étape à savoir l'**analyse** des enseignements constitue sans nul doute une des plus importantes du cycle. Entamée il y a plusieurs années, elle a fait l'objet de modifications régulières pour identifier les diverses sections de chaque chapitre et au sein de chacune d'entre elles les points d'enseignements (PE). Chacun de ces 279 PE définis s'est vu attribuer une cote de priorité de 1 (moins important : savoir que cela existe) à 3 (très important : à connaître absolument). Trois niveaux de performances (CP : catégories de performances) ont été distingués. Par connaissance, il faut entendre la capacité de connaître une série de faits ou d'énoncer des définitions. Par compréhension, il faut comprendre la capacité à interpréter des symptômes, et relations de ou entre l'une ou l'autre pathologie de reproduction, obstétricale ou mammaire ou encore la capacité à extrapoler le sens d'un message, à en saisir la nature et la signification profonde. Par application, il faut comprendre la capacité de mettre en œuvre l'un ou l'autre traitement individuel ou collectif. Ont ainsi une fois définis, ces PE et CP ont été croisés en vue de déterminer des binômes [PExCP]. C'est ainsi que respectivement, 92, 37 et 20 % des PE ont fait l'objet d'un CP dit de connaissance, de compréhension et d'application.
La seconde étape concerne la définition du **design** du test envisagé à savoir les modalités du questionnement en fonction de l'objectif cognitif poursuivi (exemple taxonomique de Bloom). Cette étape implique aussi de définir la proportion de questions en fonction du binôme [PExCP] défini à l'étape 1. Ainsi, dans le cadre de cette étude, nous avons été attentifs à ce que les tests de novembre et de janvier comportent respectivement 20 et 10 questions de connaissance, 15 et 20 questions de compréhension et 5 et 10 questions d'application.
La troisième étape concerne les **questions.** Pour la rédaction des questions QCM-SGI, nous nous référons aux principes définis par Leclercq (1986). La plateforme rend possible la rédaction en ligne des questions pour les CP définis. Le professeur et ses assistants ont donc ainsi la possibilité d'élargir la base de questions du cours. S'en suit une étape de validation avec l'ensemble des rédacteurs. Elle peut entraîner le rejet de la question si cette dernière apparaît équivoque. Une vérification du respect des règles de conception des questions est faite en même temps. Il y a donc un contrôle *à priori* de la qualité des questions.
La quatrième étape vise à donner aux apprenants la possibilité de s'**entraîner** au type de test organisé. Dans notre cas cet entraînement est assuré par des tests formatifs mis en ligne au moyen de la plateforme WebCT. Ces questions formatives sont pour la plupart les questions rédigées et utilisées pour les tests certificatifs des années précédentes. En 2007-2008, 31 tests formatifs

comprenant 5 à 15 questions selon les chapitres, ont été mis en ligne à la disposition des étudiants via la plateforme WebCT. Ils ont été réalisés 249 fois en moyenne (2 à 340) soit 1,1 fois par étudiant.
La cinquième étape consiste en l'organisation proprement dite du **test**. Les épreuves certificatives se déroulent dans plusieurs amphis de manière à éviter autant que possible toute fraude. Les apprenants reçoivent l'une des 4 formes du questionnaire de 40 questions générées et mises en forme automatiquement par la plateforme ExAMS, les consignes définissant notamment les SGI, le barème des tarifs, le formulom (formulaire à lecture optique de marque) de réponse et une feuille de justification des réponses. Ils sont informés que le correcteur ne lirait que les commentaires concernant les réponses incorrectes. La justification ne peut donc QUE bénéficier à l'apprenant. Lors du test de janvier, 139 apprenants ont justifié 6,4 réponses en moyenne (1 à 21).
La sixième étape est la **correction** du test. Les réponses des apprenants consignées sur des formuloms sont lues à la lecture optique de marques puis traitées par ExAMS. Notre barème de cotation est resté inchangé depuis plusieurs années soit : + 1 point attribué en cas de réponse correcte, 0 point en cas d'omission, - 1/7ème des points en cas de réponse erronée (selon l'application de la formule 1/n-1, n représentant le nombre de propositions soit 8 dans le cas présent). Le test fait l'objet d'une vérification de sa cohérence interne. Cette vérification permet de valider *à posteriori* la qualité des questions. La correction concerne également la validation éventuelle des justifications apportées par les apprenants. Le cas échéant, une nouvelle correction du test est effectuée après ajustement. En novembre et janvier respectivement 2 et 3 questions ont fait l'objet d'une valorisation de proposition supplémentaire. De même, respectivement 9 et 36 justifications ont été acceptées pour valider les questions concernées. Une seule question a été supprimée sur l'ensemble des deux tests.
La septième étape a pour objectif de donner aux apprenants des **feedbacks** sur leur réussite ou échec. Les réponses correctes aux questions et leur justification ont été communiquées aux étudiants via la plateforme Web CT dès la fin de l'examen. Les résultats individuels finaux de chaque étudiant et les résultats moyens par catégories de performances (Figure 1) ont été communiqués à l'ensemble du groupe trois semaines après la seconde évaluation.
La dernière étape dite de **régulation** permet d'améliorer le processus de construction du test après récolte et analyse des avis des utilisateurs. Dans notre cas, cette étape consiste à supprimer de notre banque de questions une question. Elle consiste également à adresser via la plateforme WebCT un questionnaire dit de satisfaction aux apprenants. Ce questionnaire concerne l'organisation des enseignements et de leurs évaluations. Il comporte des questions à réponses ouvertes sur les points forts et faibles du système mis en place. Nous en identifierons deux plus spécifiques aux évaluations proposées. Question 1 : Apport des évaluations formatives à l'apprentissage **:** Les évaluations formatives en ligne (ou sur papier ...) m'ont aidé à assimiler la matière. Pour votre information, la majorité des questions formatives sont des questions présentées lors des tests certificatifs précédents. Question 2 : Justifications des réponses aux questions : Les justifications aux questions des tests formatifs et certificatifs m'ont aidé à mieux comprendre mes erreurs et donc à améliorer mon apprentissage. Les distributions des réponses obtenues à ces deux questions (150 /223) sont présentées dans les tableaux 1 et 2.

## 3. Discussion

De 2001 à 2006, nous avons eu recours de manière informelle à un modèle global qualitatif de gestion d'évaluations basées sur des QCM-SGI. En 2007, une impulsion nouvelle a été donnée à notre approche par le recours à un système informatisé de gestion de cette approche qualitative qu'est la plateforme ExAMS. Ce faisant, nous avons pu largement modifié notre rapport aux activités d'évaluation (Jorro 2006). Volontairement nous sommes éloignés de pratiques de tâtonnement par imitation pour nous inscrire dans une approche plus rigoureuse et valoriser ainsi autant que faire se pouvait, les notions de validité, fidélité, sensibilité, diagnosticité, praticabilité, équité, communicabilité et authenticité d'une évaluation. Nous nous sommes ainsi inscrits en droite ligne dans les impératifs de qualité souhaités par la Conférence des Ministres européens chargés de l'Enseignement Supérieur, (http://www.fage.asso.fr/communique_bergen.php). Cette démarche n'a cependant pu être possible que par une remise en question et une adaptation régulière au cours de ces dernières années de nos objectifs d'enseignements. Cette étape préliminaire nous a grandement facilité la mise en place d'une approche par objectifs d'évaluation et a renforcé nos compétences méthodologiques. L'outil ExAMS a apporté dans ce contexte une aide incontournable. Il ne nous est pas possible à ce jour de présenter concrètement l'impact de ce changement de paradigme (objectifs d'évaluation vs objectifs d'enseignement) sur la maîtrise des apprentissages par nos apprenants. Plus encore que les années précédentes, nous avons informé nos apprenants de notre démarche et des outils utilisés. Le cycle de gestion qualitative a fait l'objet d'une présentation en présentiel. Attendu l'intérêt manifesté par les apprenants, l'expérience sera bien entendu renouvelée au cours des prochaines années. Ce faisant, nous nous inscrivons dans une démarche de clarification de l'évaluation jugée indispensable puisqu'en effet présenter une note est une chose et la manière dont elle a été obtenue en est une autre (Dauvisis 2006). Cette approche est un premier exemple du renforcement possible des compétences sémiotiques possibles des enseignants.
Réussir et/ou apprendre : that's the question que se pose bien évidemment tout apprenant. La réponse est classique : dans la majorité des cas, les apprenants vont privilégier la réussite et mettront en place une stratégie d'apprentissage qui peut y contribuer.
Il nous incombe donc (1) de recourir plus systématiquement à des objectifs cognitifs de plus haut niveau que ceux qui relèvent de la simple connaissance, (2) de forcer « malgré eux » les apprenants à développer leurs capacités bien réelles à comprendre malgré tout, à appliquer voir à évaluer et (3) à mettre en place un système de feedback qui leur permet de mieux appréhender les causes de leur échec éventuel. Dans ce contexte, l'évaluation peut se concevoir comme une valorisation des possibles et le debriefing d'une évaluation prend tout son sens (Jorro 2005). La plateforme ExAMS offre de réelles possibilités de gérer de manière optimale ces trois aspects, de manière collective mais aussi individuelle. Indirectement, elle permet par ailleurs de changer l'image de l'évaluateur. Il a la possibilité de devenir « l'ami critique » qui sait allier tout à la fois la bienveillance et l'exigence, la bienveillance au travers de la collecte d'avis et l'exigence au travers d'une communication et de la valorisation du potentiel de l'évalué (MacBeath 1998). On le constate, lentement mais sûrement on se déplace du paradigme de « l'enseignant-qui-enseigne-et-qui-évalue » vers un autre paradigme de « l'élève-qui-apprend-y-compris-en-participant-au-processus-de-l'évaluation » (Figari 2006).

L'expérience décrite est en marche. Elle s'est avérée intéressante car elle a ouvert de nouvelles perspectives. Au nombre de celles-ci, nous mentionnerons pour la prochaine année académique 2008-2009

- le recours plus systématique aux degrés de certitude pour procéder à l'évaluation d'un niveau supérieur d'objectifs cognitifs. La plateforme ExAMS présente à cet égard un avantage par rapport à la platefome WebCT puisqu'elle offre la possibilité de recourir à ces cœfficients. Les apprenants auront donc l'occasion de s'y entraîner par des tests formatifs.
- la possibilité pour les apprenants d'avoir accès via la plateforme ExAMS à un feedback personnalisé de leurs tests. Cette possibilité sera de nature à développer voire à renforcer leurs attitudes réflexives à l'égard de leurs résultats.
- Le souhait de voir les apprenants participer à leurs évaluations en leur donnant la possibilité de rédiger des questions. Nous devrons résoudre la valorisation de cette activité qui peut être indirectement d'apprentissage des contenus.
- L'idée de noter différemment les questions de catégories de performances différentes.

## 4. Bibliographie


- BLOOM B. (1968) Learning for mastery. Evaluation comment, 1, 2, 34-42.
- CASTAIGNE JL, GILLES JL, HANZEN CH (2001): les Stratégies de réussite dans l'Enseignement supérieur. Application du cycle gestion qualité SMART des tests pédagogiques au cours d'obstétrique et de pathologie de la reproduction des ruminants, équidés et porcs. In Actes du *18ème congrès de l'AIPU* (Dakar).
- DAUVISIS MC. (2006). L'instrumentalisation en évaluation. Mesure et Evaluation en Education, 29, 1, 45-66.
- GILLES J.-L. et 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. (2002). Qualité spectrale des tests standardisés universitaires : Mise au point d'indices édumétriques d'analyse de la qualité spectrale des évaluations des acquis des étudiants universitaires et application aux épreuves MOHICAN check up '99. Thèse de doctorat, Université de Liège
- FIGARI G. (2006) ; L'activité évaluative entre cognition et réponse sociale : nouveaux défis pour les évaluateurs. Mesure et Evaluation en Education, 29, 1, 5-18.
- JORRO A. (2006). Devenir un ami critique ; Avec quelles compétences et quels gestes professionnels ? Mesure et Evaluation en Education, 29, 1, 31-44.
- JORRO A. (2005). L'éthique de l'évaluateur. Conférence au séminaire 'Audit d'établissement ». Poitiers ESEN
- KARRAKER, R. J. (1967). Knowledge of results and incorrect recall of plausible multiple-choice alternatives. Journal of Educational Psychology, 58, 11-14.
- LECLERCQ D. (1986). La conception des questions à choix multiple, Bruxelles, Ed. Labor.
- LECLERCQ D. *et al.* (1993). 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. (1999). Édumétrie et docimologie, Liège : STE-ULG, 1999.
- MAC BEATH J. (1998). I didn't know he was ill : the role of the critical friend. In L.Stoll and JK SMyers (eds), No quick fixes : perspectives in school in difficulty. London, Falmer Press
- WOOD R. (1977). Multiple choice : A state of the art report. In Chopin & Postlewaite (eds), Evaluation in Education International Progress. Oxford : Pergamon


## 5. Tableaux et figures

Figure 1 : Comparaison des résultats moyens par catégorie de performances (/20)

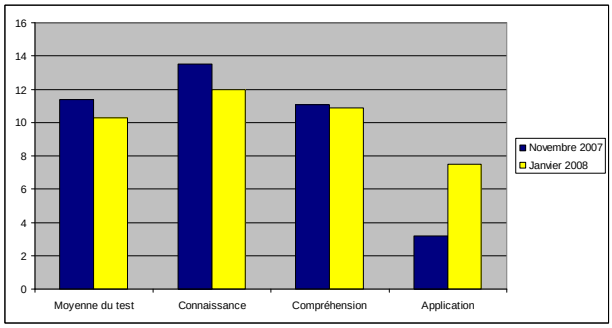

Tableau 1 : distribution des réponses à la question 1

| Réponse | Distribution de fréquence |
|---|---|
| 1. Je n'ai pas d'avis car j'ai fait moins de 50 % des tests formatifs | **10 (6,3%)** |
| 2. Pas du tout d'accord | **6 (3,8%)** |
| 3. Pas d'accord | **18 (11,3%)** |
| 4. D'accord | **92 (57,9%)** |
| 5. Tout à fait d'accord | **33 (20,8%)** |

Tableau 2 : distribution des réponses à la question 2

| Réponse | Distribution de fréquence |
|---|---|
| 1. Sans avis, je ne les ai pas consultées | **2 (1,3%)** |
| 2. Pas du tout d'accord | **2 (1,3%)** |
| 3. Pas d'accord | **10 (6,3%)** |
| 4. D'accord | **81 (50,9%)** |
| 5. Tout à fait d'accord | **64 (40,3%)** |