

# UTILISATION DU CYCLE SMART DE GESTION QUALITE DES EVALUATIONS STANDARDISEES DANS LE CONTEXTE D'UNE HAUTE ECOLE : REGARD CRITIQUE EN TERMES DE VALIDITE, FIDELITE, SENSIBILITE DES MESURES, DIAGNOSTICITE, PRATICABILITE, EQUITE, COMMUNICABILITE ET AUTHENTICITE

JEAN-LUC GILLES<sup>1</sup> ET VINCIANE LOVINFOSSE<sup>2</sup>

<sup>1</sup> Chargé de recherches, directeur du Système Méthodologique d'Aide à la Réalisation de Tests (SMART) de l'Université de Liège - Belgique

<sup>2</sup> Professeur à la Haute Ecole Rennequin Sualem de la Province de Liège - Belgique

## ***Résumé :***

Sur base d'exemples de réalisations d'évaluations standardisées des acquis de grands groupes d'étudiants, les huit étapes du cycle de gestion de la qualité des tests conçus avec le Système Méthodologique d'Aide à la Réalisation de Tests (SMART, ULg) sont successivement analysées. Cette analyse est exposée dans une double perspective : la perspective temporelle (situation présente et projection future) et la perspective pratique des deux protagonistes (l'enseignant concepteur et évaluateur, et le service de soutien logistique et méthodologique). Du point de vue des critères qualitatifs (validité, fidélité, sensibilité des mesures, diagnosticité, praticabilité, équité, communicabilité et authenticité) la situation actuelle est décortiquée en utilisant les deux perspectives temporelle et pratique, pour en déduire des pistes d'amélioration dans une quête de l'excellence.

Le contexte est basé sur une expérience avec le SMART dans le cadre d'une évaluation certificative (examen de mi-année) pour des futur(e)s gradué(e)s en Informatique d'une Haute Ecole (première année de graduat en Informatique Industrielle, 299 étudiants).

Suite à l'analyse qualitative des différentes étapes de gestion des tests avec l'aide du SMART, des principes et conseils méthodologiques en sont déduits pour être utilisables comme règles de conception d'évaluations standardisées de qualité pour les grands groupes d'étudiants de l'enseignement supérieur.

## 1. CONTEXTE

Les processus de formation se complexifient et font de plus en plus appel aux Technologies de l'Information et de la Communication (TIC). Souvent très complexes, impliquant une multitude d'opérations techniques, ces processus quand ils s'insèrent dans un contexte pédagogique sont dans la plupart des cas pris en charge par de nouveaux acteurs dans le paysage de l'enseignement supérieur : des équipes spécialisées qui proposent leur soutien aux enseignants. Beaucoup d'institutions universitaires font appel à de telles équipes, en particulier lorsqu'il s'agit de fournir un appui pédagogique aux professeurs, notamment à ceux qui doivent faire face à de grands groupes d'étudiants (parfois plus de 600) car ceux là ne peuvent plus gérer à eux seuls tous les problèmes de la sphère de leurs activités liées aux prestations de cours et aux évaluations.

Dans l'enseignement supérieur, pratiquement toutes les grandes universités se dotent aujourd'hui de cellules spécialisées dont la mission est d'aider les enseignants. Dans le domaine de l'évaluation standardisée des grands groupes d'étudiants, le Système Méthodologique d'Aide à la Réalisation de Tests (SMART) constitue une réponse de l'Université de Liège (ULg – Belgique) au besoin d'évaluer rapidement et de façon fiable de grands auditoires. Grâce au SMART, les enseignants de l'ULg, mais aussi d'autres établissements d'enseignement supérieur, notamment certains professeurs de la Haute Ecole Rennequin Sualem, peuvent recourir à des moyens techniques modernes pour procéder à l'évaluation des acquis des grands groupes (souvent dans le 1<sup>er</sup> cycle d'étude, là où les étudiants sont les plus nombreux). Le SMART propose aussi des services aux étudiants en ayant recours aux TIC, par exemple la consultation via Internet des feedbacks individualisés de leurs performances cognitives et métacognitives après une évaluation, la réalisation d'entraînements aux procédures d'examen en amphithéâtre à l'aide de boîtiers de vote électronique ou sur Internet.

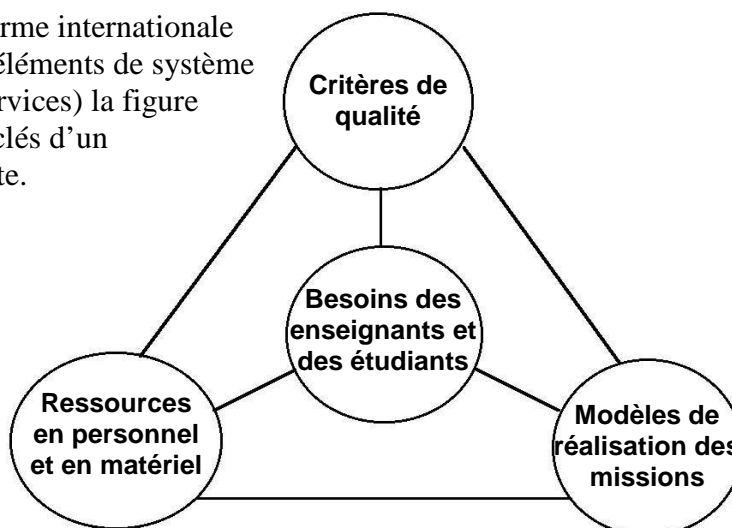
Selon Gilles (2002), la qualité des services des cellules spécialisées dont la mission est d'aider les enseignants repose sur quatre facteurs clés qui peuvent être symbolisés de la façon suivante.

Inspirée des recommandations de la norme internationale ISO 9004-2<sup>1</sup> (Gestion de la qualité et éléments de système qualité - Lignes directrices pour les services) la figure ci-contre met en évidence les facteurs clés d'un système qualité adapté à notre contexte.

Au centre : les besoins des enseignants et des étudiants, destinataires des services proposés et points de convergence des autres facteurs clés.

A la base, à gauche, les ressources en personnel et en matériel nécessaires pour la mise en œuvre du processus de réalisation des épreuves.

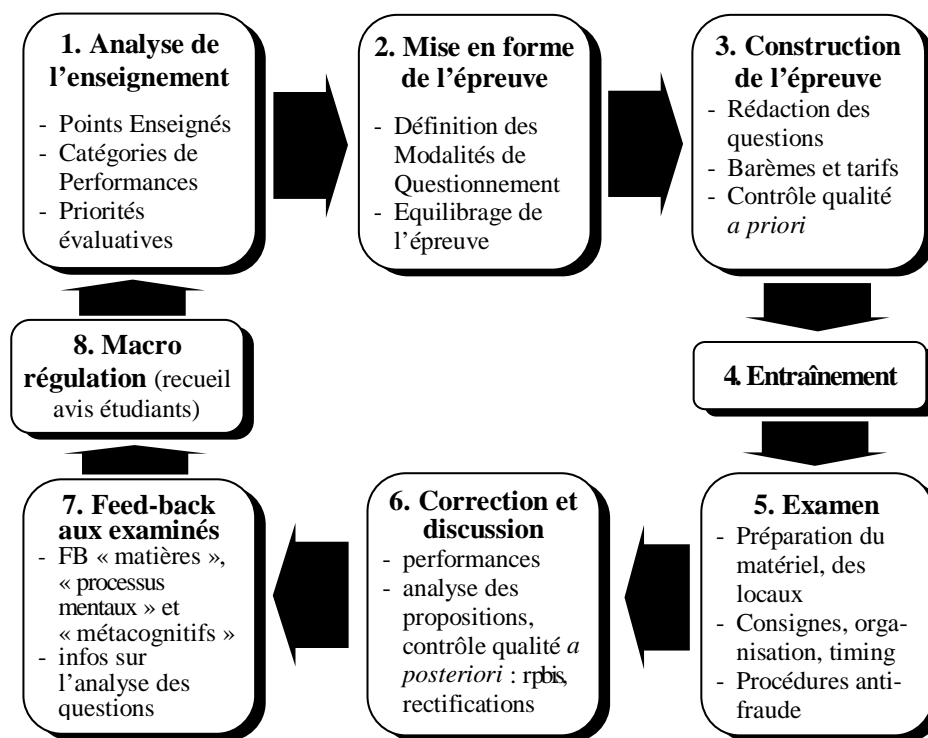
A droite, les modèles de réalisation des missions, par exemple le cycle en spirale de qualité de construction des examens qui sera détaillé plus loin. En haut du schéma nous avons placé les critères de qualité, ces facteurs sont cruciaux pour guider nos actions et nos contrôles de qualité.



<sup>1</sup> *Gestion de la qualité et éléments de système qualité – Partie 2 : Lignes directrices pour les services*, Organisation internationale de normalisation, Case postale 56, CH-1211 Genève 20, Suisse. Numéro de référence : ISO 9004-2 :1991(F). Première édition 1991-08-01, corrigée et réimprimée 1993-05-01.

## 2. LE MODÈLE MÉTHODOLOGIQUE PROPOSÉ PAR LE SMART

Le modèle méthodologique proposé par le SMART facilite la maîtrise et le contrôle de la qualité des tests. Il s'agit d'un cycle en 8 étapes : [1] analyse des objectifs, [2] mise en forme de l'évaluation, [3] construction de l'épreuve, [4] entraînement, [5] examen, [6] correction, [7] feedback aux examinés et [8] macro-régulation (Gilles et Leclercq, 1995 ; Gilles, 2002). La figure ci-dessous synthétise et schématise ces huit étapes.



[1] A l'étape de l'**analyse de l'enseignement**, la validité des tests est assurée à l'aide de procédures permettant d'identifier et de mettre en correspondance les points à évaluer et les niveaux taxonomiques (en référence à la taxonomie de Bloom) attendus chez les formés. C'est donc dès cette première étape que les niveaux taxonomiques (connaissance, compréhension, application, analyse, synthèse et évaluation) seront mis en correspondance avec les contenus de formation.

[2] Lors de la **mise en forme de l'épreuve** des modalités de questionnement sont définies en fonction des points à évaluer et des niveaux taxonomiques identifiés à l'étape précédente.

[3] L'étape de **construction de l'épreuve** permet ensuite de concrétiser en une série d'items de testing le schéma d'épreuve établi lors des deux premières étapes. Des contrôles qualité *a priori* des questions auront lieu (utilisation de règles de rédaction, validation par des experts, ...).

[4] Un **entraînement** systématique permettra de familiariser les formés avec les procédures de test planifiées aux étapes précédentes.

[5] Lors de l'étape de l'**examen** des procédures permettant une définition claire, précise et exhaustive des modalités d'utilisation du test et de ses consignes sont définies, des procédures anti-fraude sont également mises en place.

[6] Lors de l'étape de **correction et discussion**, des procédures de contrôles qualité *a posteriori* des questions sont proposées et ont recours à des indices classiques (indices de cohérences internes). Des rectifications des résultats des épreuves peuvent être proposées suite aux contrôles qualité.

[7] La création de rétro-informations à l'étape des **feedbacks aux examinés** permet aux formés

d'identifier clairement leurs points forts et leurs faiblesses.

[8] Enfin, l'étape de **macro-régulation** permet d'améliorer le processus de construction du test après récolte et analyse des avis des utilisateurs.

Dans ce contexte, l'équipe du SMART tient à jour un dossier de suivi de la réalisation de chaque test. Ces dossiers contiennent les documents attestant des contrôles qualité effectués aux différentes étapes du cycle et permettent le traçage et la validation du suivi de la qualité des tests.

### 3. LES CRITÈRE DE QUALITÉ UTILISÉS PAR LE SMART

Les procédures et les dispositifs d'ingénierie docimologiques proposés par le SMART rencontrent une série de critères de qualité (Gilles, 2002) :

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

### 4. PROCÉDURES ET DISPOSITIFS TECHNOLOGIQUES

Actuellement, plus d'une trentaine de procédures sont utilisées lors des étapes du cycle SMART en vue d'améliorer la qualité des épreuves (les cadres placés près des différentes étapes sur la figure suivante).

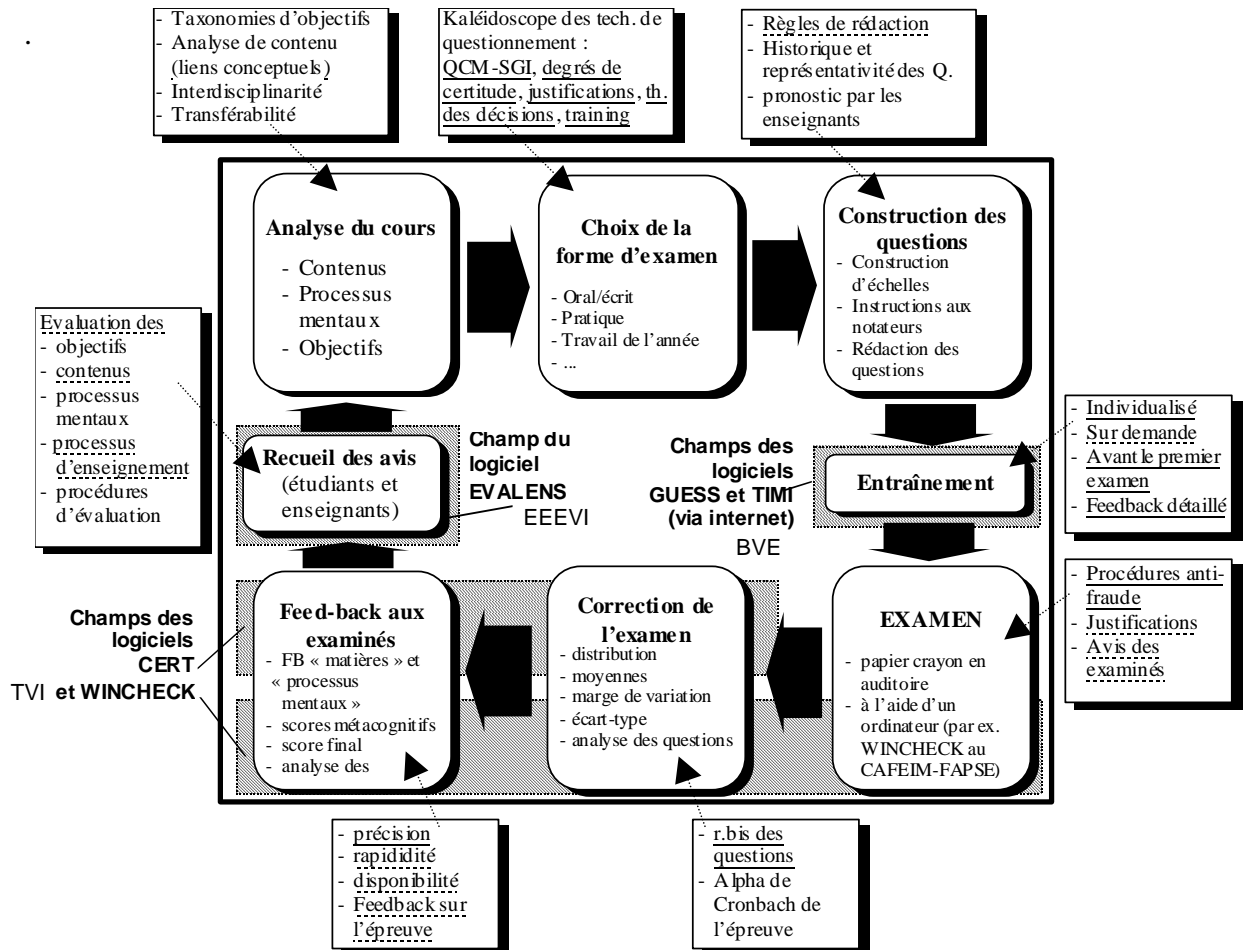
Une série de dispositifs technologiques qui couvrent certaines étapes du cycle de construction des épreuves (zones en hachuré sur le schéma) nous permettent de répondre aux demandes (actuellement en forte croissance) des enseignants. Trois types de technologies sont utilisées et souvent en couplage les unes avec les autres :

- Une chaîne complète de lecture optique de marques permettant d'utiliser différents formulons (FORMulaires pour la Lecture Optique de Marques). Ces feuilles de réponse ont été conçues et programmées par le SMART pour faciliter les évaluations standardisées des acquis des étudiants. Des logiciels, dont certains ont été créés sur mesure, sont prévus pour traiter et corriger les tests, voire pour permettre l'évaluation interactive des acquis des étudiants (CERT, WINCHECK).
- Les Boîtiers de Votes Electroniques (BVE) permettent à une centaine d'étudiants (ou à 100 groupes d'étudiants) réunis dans un même amphithéâtre de fournir des réponses codées de zéro à 9 (donc de répondre à des QCM). La ventilation des résultats peut être

*Utilisation du cycle SMART de gestion qualité des évaluations standardisées dans le contexte d'une Haute Ecole : regard critique en termes de validité, fidélité, sensibilité des mesures, diagnosticité, praticabilité, équité, communicabilité et authenticité*

immédiatement projetée sur l'écran dans la salle de cours, ce qui permet un feedback immédiat sur la qualité des réponses fournies par le groupe.

- Des serveurs web utilisés notamment pour envoyer des feedbacks via internet après un test ayant recours à la lecture optique de marques, pour entraîner les étudiants à des procédures d'évaluation sophistiquées, par exemple les pourcentages de certitude (GUESS), ou pour permettre de passer des examens d'années précédentes « à blanc » (TIMI), ou encore pour recueillir les avis des étudiants à propos des examens qu'ils ont subi (EEEVI).



Ces dispositifs et ces procédures sont utilisés depuis plusieurs années en routine par le SMART et ont fait l'objet de nombreuses publications et communications où elles sont décrites en détail (Detroz et Noël, 2002 ; Gilles & Leclercq, 1995 ; Gilles, 1997, 1998, 2002, 2004 ; Castaigne, Gilles et Hansen, 2001 ; Gilles, Bourguignon et Detroz, 2000).

## 5. ANALYSE DES ÉTAPES EN FONCTION DES CRITÈRES DE QUALITÉ : LE CAS DU COURS DE SYSTÈME DE GESTION DE BASES DE DONNÉES

Dans ce qui suit nous allons analyser chacune des étapes (perspective temporelle du processus de construction) du point de vue des critères qualitatifs (validité, fidélité, sensibilité des mesures, diagnosticité, praticabilité, équité, communicabilité et authenticité). Cette analyse a lieu après une première utilisation par une enseignante du modèle méthodologique proposé par le SMART. L'enseignante, Madame Vinciane Lovinfosse, est professeur d'informatique en première année du graduat en informatique et système de la Haute Ecole de la Province de Liège (HEPL - Rennequin Sualem). L'évaluation certificative analysée concernait un cours théorique d'informatique intitulé *Systèmes de gestion de bases de données* donné à une section comportant environ 300 étudiants.

*Utilisation du cycle SMART de gestion qualité des évaluations standardisées dans le contexte d'une Haute Ecole : regard critique en termes de validité, fidélité, sensibilité des mesures, diagnosticité, praticabilité, équité, communicabilité et authenticité*

Notre intention est d'aboutir après analyse à une série de pistes d'améliorations dans une quête de l'excellence

En ce qui concerne la première étape d'*analyse de l'enseignement*, c'est avant tout le critère de validité qui entre en jeu. Pour s'assurer de la couverture de l'ensemble de la matière, il conviendrait de dresser la liste des points enseignés et de déterminer les priorités quant à la matière qui devra être évaluée. Pour ce qui est de permettre des inférences solides en ce qui concerne les processus mentaux qui seront sollicités chez les étudiants, il conviendrait de clarifier les catégories de performances attendues (actuellement la connaissance de mémoire et la compréhension au sens défini par Bloom dans sa célèbre taxonomie). Cette première étape permet aussi de réunir une série d'informations en vue d'effectuer ultérieurement un diagnostic des points maîtrisés de ceux qui ne le sont pas et pareillement en ce qui concerne les processus mentaux envisagés. Enfin, l'analyse de la matière à évaluer doit permettre de réunir d'autres informations qui permettront de construire des questions en rapport avec des situations qui ont du sens pour les étudiants et qui sont pertinentes par rapport au monde réel, ce qui renforce le critère d'authenticité.

A la seconde étape de *mise en forme de l'épreuve* qui déterminera le choix des différents types de modalités de questionnements, le critère de validité continue de jouer. Actuellement, dans le cadre de son cours d'informatique, avant de construire ses questions, l'enseignante parcourt ses notes et détermine en fonction des catégories de performances définies à l'étape précédente quelles modalités de questionnement pourraient être utilisées. Il nous semble que ce travail de clarification systématique à propos des modalités de questionnement devrait aussi contribuer à améliorer la consistance intra juge, donc la fidélité de l'épreuve. La fidélité est également améliorée par le choix de techniques de questionnement ayant recours aux questions à choix multiple étant donné la possibilité d'automatiser la correction et d'éviter ainsi les biais classiques liés à l'introduction de la subjectivité des correcteurs dans l'analyse des réponses. Le recours aux QCM permet aussi de poser de nombreuses questions et donc de mieux couvrir l'ensemble de la matière, ce qui renforce encore le critère de validité. Le choix par l'enseignante d'une utilisation combinée de QCM avec pourcentages de certitude permet aussi d'améliorer la sensibilité des mesures (voir étape suivante). L'examen proposé comporte aussi quelques questions ouvertes qui permettent d'évaluer certains objectifs qui ne peuvent pas l'être avec des QCM, par exemple la capacité à faire preuve de créativité dans le cadre de résolutions de problèmes informatiques et de poser des questions qui peuvent contribuer à une amélioration de l'authenticité. La communication d'explications à propos des différents types de questionnements auxquels les étudiants seront confrontés devrait permettre d'améliorer le critère de communicabilité. Le fait de choisir des modalités de questionnement qui permettent de poser les mêmes questions à tous les étudiants renforce aussi l'équité de l'examen.

Dans le cadre de la troisième étape de *construction des questions*, l'enseignante améliore la validité et la fidélité de l'épreuve en soumettant les questions du test à un groupe de collègues. Les critiques de ces derniers permettent un contrôle qualité *a priori* (avant que l'épreuve n'ait eu lieu) ce qui réduit les risques d'items mal formulés ou non pertinents. En fait, c'est un véritable travail d'équipe qui est effectué à cette étape et le regard extérieur des collègues peut contribuer à une amélioration de l'authenticité des questions. Le barème des tarifs utilisés est choisi collégalement et dans le cadre de cette épreuve d'informatique il a été décidé d'utiliser les pourcentages de certitude avec une tarification classique établie par Leclercq (1993), ce qui renforce le critère de sensibilité des mesures car cette technique permet de dépasser une vision binaire des réponses « *c'est soit vrai, soit faux...* » en autorisant des nuances subtiles dans un spectre de réponses. Par exemple, à l'aide des pourcentages de certitude il devient possible de différencier *une réponse correcte accompagnée d'un degré de certitude faible*, ce qui revient pratiquement à de l'ignorance, d'*une réponse correcte accompagnée d'une certitude très élevée* ce qui correspond à un état de connaissance assurée. Pour plus de détail sur ces nuances spectrales des réponses des étudiants, le lecteur consultera les travaux

de Jans et Leclercq (1999). Parallèlement, lors de cette troisième étape l'équipe définit les exigences attendues et les tarifs associés aux questions à réponse ouverte ce qui permet aussi d'en améliorer la fidélité.

A la quatrième étape d'*entraînement des étudiants*, l'enseignante procède à une information des étudiants à la fin de plusieurs séances de cours où elle explique ou rappelle les consignes et propose quelques questions représentatives. Ce type d'entraînement renforce les critères de qualité de communicabilité et aussi d'équité dans la mesure où l'information sur les modalités d'examen ne sont plus l'apanage des seuls étudiants qui ont échoué l'année précédente et qui recommencent leur année d'étude. Nous pensons que l'entraînement permet aussi d'améliorer la fidélité des résultats moins entachés de « *bruits* », au sens de la théorie de la communication, lorsque les étudiants sont entraînés que lorsqu'ils découvrent des modalités de questionnements sophistiquées le jour même de l'examen. Enfin, un entraînement à l'utilisation des degrés de certitude est disponible en ligne sur le site du SMART. Il s'agit du logiciel *GUESS*<sup>2</sup> qui permet en peu de temps de fournir de nombreuses réponses accompagnées d'un pourcentage de certitude, le but du jeu étant de deviner les lettres d'un texte. L'interface fournit un feedback immédiat sur la qualité de ses auto-évaluations ce qui permet d'ajuster la façon dont on utilise les certitudes. L'utilisateur peut ainsi gérer les probabilités qu'il fournit en veillant à ne pas se sur ou sous-estimer.

Pour la cinquième étape, la *mise en œuvre de l'examen*, la fidélité est améliorée par la concertation préalable de l'équipe des enseignants à propos du timing et des consignes de l'épreuve. Le souci d'équité est particulièrement présent à cette étape : l'enseignante veille sélectionner des locaux identiques pour accueillir les différents groupes étudiants. Des procédures anti-fraude sont mises en place : les étudiants sont répartis dans les locaux de façon suffisamment espacée, de plus, ils ont tous les mêmes questions, mais pas dans le même ordre que leur voisin. La communicabilité est aussi améliorée par la mise à disposition d'une feuille de consigne détaillée, et, au début de l'examen, une check-list est passée en revue avec les étudiants pour vérifier point par point les conditions matérielles de réalisation de l'épreuve. L'enseignante rappelle aussi les consignes de bonne utilisation des feuilles spécialement prévues pour la lecture optique des réponses des étudiants aux QCM. Dans le contexte de cette enseignante, la facilité de mise en œuvre de ce système est par ailleurs à mettre en relation avec le critère de praticabilité.

Après l'examen proprement dit, vient la sixième étape de *correction et discussion des résultats*. La fidélité et l'équité sont garanties dans le cadre de l'automatisation de la correction via le système de lecture optique de marques qui permet aussi de renforcer le critère de praticabilité à cette étape. Par ailleurs, l'analyse de la qualité des questions à l'aide d'indices comme les corrélations point bisériales (*rpbis*) permet de renforcer la fidélité des résultats, certaines QCM où des problèmes sont détectés à l'aide de ces analyses pouvant être supprimées ou rectifiées (changement de réponse correcte ou/et valorisation d'un distracteur), dans une certaine mesure, on contribue aussi ainsi à améliorer la validité de l'examen. Ce recul objectif sur la qualité des questions renforce aussi l'équité dans la mesure où les rectifications ont des répercussions sur tous les étudiants. C'est aussi à cette étape que sont effectués les traitements en vue d'établir les diagnostics individualisés des compétences des étudiants.

Lors de la septième étape des *feedbacks aux étudiants* c'est plus particulièrement le critère de qualité de communicabilité qui est rencontré ainsi que les critères d'équité en terme d'accès à l'information, de diagnosticité et de praticabilité par le fait de l'utilisation d'une procédure qui permet aux étudiants de consulter le feedback détaillé et individualisé de leur résultat sur le web. Les procédures proposées par le SMART libèrent l'enseignante des tâches complexes liées à la mise

---

<sup>2</sup> Le logiciel GUESS est utilisable à l'adresse : <http://www.smart.ulg.ac.be/smart/ee/guess/>

sur l'internet de ces feedbacks, en ce sens le critère de praticabilité est particulièrement rencontré à cette étape.

Enfin, dans le cadre de la huitième et dernière étape de **macro-régulation**, l'enseignante envisage l'utilisation d'un formulaire de satisfaction. Actuellement, dans une certaine mesure, se sont les critères d'équité et de communicabilité qui sont renforcés à cette étape. Des discussions sont possibles, l'enseignante donne un « *droit à la parole* » à chaque étudiant, ces derniers étant par ailleurs considérés comme des partenaires avec qui le dialogue est possible dans le processus de construction et de gestion de la qualité des tests.

En synthèse, la figure ci-dessous montre par estimation dans quelle mesure l'accent est mis actuellement sur l'un ou l'autre critère de qualité aux différentes étapes, et ce, dans le contexte de l'examen du cours *système de gestion de bases de données*. La présence ou l'absence d'une pastille dans les cases indique la prise en compte ou non d'un critère de qualité précis à une étape donnée dans le cadre de la construction de l'épreuve d'informatique qui a été analysée. La grosseur d'une pastille donne notre estimation de l'importance accordée actuellement au critère.

Nous avons placé des signes « \* » dans la dernière ligne du tableau car en fonction des informations qui seront recueillies à l'étape de macro-régulation du processus de construction de l'évaluation (avis des étudiants, des collègues, auto-évaluations de l'enseignante, ...) une série de décisions liées aux différents critères de qualité pourraient être prises en vue d'améliorer l'épreuve. C'est d'ailleurs à ce type de réflexion que nous allons nous livrer dans le cadre du point suivant « pistes et propositions d'améliorations ».

	Validité	Fidélité	Sensibilité	Diagnosticité	Praticabilité	Équité	Communicabilité	Authenticité
1. Analyse enseignement/formation	●			●				●
2. Mise en forme de l'épreuve	●	●	●	●		●	●	●
3. Construction des questions	●	●	●					●
4. Entraînement des étudiants		●			●	●	●	
5. Mise en œuvre de l'évaluation		●			●	●	●	
6. Correction et discussion	●	●		●	●	●		
7. Feedbacks aux étudiants				●	●	●	●	
8. Macro-régulation du processus	*	*	*	*	*	●	●	*



## 6. PISTES ET PROPOSITIONS D'AMÉLIORATIONS

Pour chacune des étapes, nous nous sommes livrés à une analyse de ce qui pourrait être amélioré tant du côté de l'enseignante que du côté du SMART.

En ce qui concerne l'étape d'**analyse de l'enseignement / formation**, du point de vue du critère de validité, l'enseignante estime qu'elle devrait encore clarifier les catégories de performances (CP) attendues chez les étudiants et appliquer la « *méthode des trinômes* » conçue par le SMART (Castaigne, Gilles et Hansen, 2001). Appliquée à cette étape, cette méthode permettrait d'aboutir à une table des spécifications où les points de matière à évaluer (PE) sont systématiquement mis en correspondance avec les catégories de performances attendues chez les étudiants. Ce travail de conception de l'épreuve serait facilité si le SMART informatisait la méthode proposée. Une explication des procédures devrait aussi être diffusée à terme. Ces outils renforceraient les qualités de validité mais aussi de praticabilité de l'épreuve. La validité pourrait aussi être mieux contrôlée si le SMART offrait des indices permettant de la quantifier. Par exemple, à l'aide d'indices de couverture de la matière dont le calcul prendrait aussi en compte les priorités accordées par l'enseignant. Du point de vue du critère de communicabilité, la liste des points à évaluer et les définitions des catégories de performances devraient être communiquées aux étudiants avant l'épreuve.

A l'aide de la table de spécification qui pourrait être créée via la méthode dite « des trinômes » l'enseignante pourrait aussi adapter de façon plus systématique les modalités de questionnement (MQ) aux catégories de performances (CP). Ce travail serait réalisé dans le cadre de la seconde étape de *mise en forme de l'épreuve*. Ici aussi, un outil informatisé proposé par le SMART faciliterait le travail des enseignants (amélioration de la praticabilité).

Lors de la troisième étape de *construction des questions*, l'enseignante compte améliorer la fidélité des questionnaires en les soumettant systématiquement à une relecture formelle de l'équipe du SMART. L'enseignante pré-testera aussi ses questions auprès de ses collègues. Cet autre contrôle qualité (qui ne nécessite pas l'intervention du SMART) pourrait encore améliorer la validité de l'épreuve. Le SMART faciliterait aussi le travail des enseignants en fournissant un système informatisé de gestion des banques de questions, et renforcerait ainsi les critères de validité et de praticabilité.

Pour l'étape d'*entraînement*, l'enseignante diffusera les examens corrigés des années précédentes. Cette procédure simple (impossible à mettre en œuvre dans le cadre de cette première utilisation) renforcera les qualités de communicabilité, de fidélité et d'équité. Le SMART améliore actuellement un outil de *testing interactif multimédia via internet* (TIMI) afin de le rendre plus convivial. TIMI permettra une mise en ligne des examens des années précédentes avec des facilités de correction et de feedbacks immédiats pour les étudiants qui utiliseront le système dans le cadre d'un entraînement. Un tel outil contribuera encore à améliorer la praticabilité de ces examens « à blanc ». La systématisation de l'entraînement pourrait aussi permettre d'envisager des modalités de questionnement plus subtiles et probablement plus sophistiquées (ce qui rend l'entraînement indispensable), dès lors il y aurait des retombées qualitatives en termes de sensibilité des mesures.

Dans le cadre de la cinquième étape de *mise en œuvre de l'examen*, l'enseignante compte prendre note des remarques des étudiants de façon systématique lors de l'examen. Cette procédure permettrait de garder la trace d'éventuels problèmes signalés par les étudiants et éventuellement liés à des ambiguïtés ou des erreurs dans les questions. Ces informations fourniraient à l'enseignante un éclairage qualitatif pour interpréter l'analyse quantitative de l'épreuve (notamment à l'aide des corrélations point bisérialles) effectuée à l'étape suivante. On renforcerait ainsi la fidélité de

l'épreuve. Le SMART de son côté pourrait améliorer la praticabilité en offrant la possibilité aux étudiants de cocher l'appartenance à une sous-section, ce qui faciliterait l'obtention de sous-scores par sous-sections. La fidélité pourrait aussi être améliorée par une feuille de consigne qui inclurait une check-list du type « *vous avez terminé, avant de rendre votre copie, vérifiez ...* ».

Pour la sixième étape de **correction et discussion des résultats**, c'est essentiellement au niveau du SMART que des améliorations pourraient être apportées. Une série d'indices supplémentaires devraient être proposés comme l'alpha de Cronbach ou le calcul de *rpbis spectraux* (Gilles, 2002) des explications sur ces différents indices devraient aussi être disponibles. Par ailleurs, une série de simulations sur les résultats pourraient être effectuées par l'enseignant lui-même. Ces mesures permettraient de renforcer la fidélité et la praticabilité.

En ce qui concerne l'étape des **feedbacks aux étudiants**, la diffusion après l'examen du questionnaire et du corrigé pourrait être améliorée en termes de communicabilité. Dans le cadre des feedbacks via internet la praticabilité du système pourrait encore être améliorée par une série de facilités qui seraient offertes par le SMART, par exemple la possibilité d'afficher avec plus de convivialité le contenu d'une question lorsque l'étudiant visionne ses résultats. Le SMART devrait aussi prévoir une procédure d'impression et d'envoi postal des feedbacks pour les étudiants qui n'ont pas la possibilité de se connecter à l'internet. Une procédure officielle de réclamation (la même pour tous les étudiants) en cas de problème, en lien avec un examen, devrait aussi être proposée sur le site du SMART. Ces deux dernières mesures renforceraient le critère d'équité.

Enfin, pour ce qui est de la huitième étape de **macro-régulation**, le SMART pourrait améliorer la praticabilité en informatisant la procédure de recueil des avis des étudiants. Cette procédure devrait inclure des items « sur mesure » définis par l'enseignant. Des procédures complémentaires qui permettraient d'évaluer au niveau du marché de l'emploi le transfert et l'intérêt des compétences apprises pourraient renforcer la validité des épreuves.

La figure ci-dessous offre une visualisation des apports des mesures que nous proposons. Par rapport aux pastilles de la figure précédente, la zone blanche symbolise une estimation de l'impact des initiatives prises du côté enseignant tandis que la zone grisée reprend une estimation de l'impact des mesures qui devront être initiées par le SMART.

	Validité	Fidélité	Sensibilité	Diagnosticité	Praticabilité	Équité	Communicabilité	Authenticité
1. Analyse enseignement/formation	●			●	○		○	●
2. Mise en forme de l'épreuve	●	●	●	●	○	●	●	●
3. Construction des questions	●	●	●		○			●
4. Entraînement des étudiants		○	○		○	○	●	
5. Mise en œuvre de l'évaluation		○			○	●	●	
6. Correction et discussion	●	○		●	○	●		
7. Feedbacks aux étudiants				●	●	○	○	
8. Macro-régulation du processus	○	*	*	*	○	●	●	*

## 7. CONCLUSIONS

Notre étude met en évidence des possibilités d'amélioration des méthodes et des dispositifs proposés par le SMART. Nos constats se limitent au contexte de la construction et de la gestion de la qualité d'une épreuve certificative d'informatique proposée à un grand groupe d'étudiants de l'enseignement supérieur.

En ce qui concerne la validité, de nouvelles initiatives devraient être prises au niveau des trois premières étapes du cycle : « 1. *Analyse de l'enseignement/formation* », « 2. *Mise en forme de l'épreuve* » et « 3. *Construction des questions* ».

La fidélité des épreuves pourrait être améliorée aux étapes « 4. *Entraînement des étudiants* », « 5. *Mise en œuvre de l'épreuve* » et « 6. *Correction discussion* ».

En ce qui concerne le critère de sensibilité des mesures, il nous semble que des possibilités d'améliorations seraient envisageables en ayant recours à des entraînements plus systématiques ayant recours des procédures plus sophistiquées d'évaluation.

Dans le cadre de cette étude nous n'avons pas « détecté » de nouvelles possibilités d'améliorer la diagnosticité de l'épreuve proposée, mais probablement ce constat limité à ce contexte bien précis (première utilisation du cycle SMART dans le contexte d'un test d'informatique) pourrait-il être remis en question à l'occasion d'autres études de ce type réalisées dans d'autres situations d'évaluation.

La praticabilité pourrait être améliorée pratiquement à toutes les étapes. Une importante recherche en cours<sup>3</sup> devant aboutir à l'informatisation d'une plate-forme électronique de construction et de gestion qualité des tests standardisés devrait beaucoup contribuer à l'amélioration du dispositif de ce point de vue « praticabilité ».

L'équité pourrait être encore renforcée par des entraînements plus systématiques via le web. La mise à disposition des examens des années précédentes à l'aide d'un programme de testing (TIMI) accessible par l'internet devrait être appréciée des étudiants en informatique concernés par cette épreuve du cours de *Systèmes de gestion de bases de données*.

Le critère de communicabilité devrait quant à lui être mieux pris en compte à la première étape d'analyse de l'enseignement par la diffusion d'informations sur les points à évaluer et les catégories de performances attendues. La systématisation des entraînements contribuerait aussi à une meilleure « communicabilité ». Enfin, il ressort également de notre étude que les feedbacks aux étudiants pourraient être améliorés de ce point de vue.

Dans le contexte de l'épreuve d'informatique qui a servi de base à cette analyse nous n'avons que peu envisagé le critère d'authenticité. Probablement que dans d'autres domaines ce critère aurait amené de nouvelles propositions d'améliorations, en particulier à l'étape de construction des questions.

---

<sup>3</sup> Il s'agit de la recherche intitulée « *electronic Construction and Quality Control in Standardized Testing (e-C&QCST)* » coordonnée par le SMART et menée en partenariat avec les HEC-Liège. Cette recherche est commanditée par la Direction Générale des Technologies, de la Recherche et de l'Energie (DGTRE) de la Région Wallonne – Belgique.

*Utilisation du cycle SMART de gestion qualité des évaluations standardisées dans le contexte d'une Haute Ecole : regard critique en termes de validité, fidélité, sensibilité des mesures, diagnosticité, praticabilité, équité, communicabilité et authenticité*

Les réflexions issues de ces analyses ont permis de déboucher sur de nouvelles pistes d'amélioration du cycle de construction des évaluations. Ceci nous incite à renouveler ce type d'étude dans d'autres contextes d'évaluation des performances des étudiants. Ces recherches devraient permettre (1) d'étudier dans quelle mesure ces avancées sont généralisables à différentes situations d'évaluation et (2) de mettre à jour de nouvelles propositions d'améliorations de nos méthodes en fonction des critères de qualité que nous avons identifiés.

\*\*\*

## BIBLIOGRAPHIE

- Castaigne, J.-L., Gilles, J.-L. et Hansen, C. (2001). Application du cycle gestion qualité SMART des tests pédagogiques au cours d'Obstétrique et de Pathologie de la Reproduction des ruminants, équidés et porcins, Communication au 18<sup>ème</sup> Congrès de l'Association Internationale de Pédagogie Universitaire (AIPU), *Les stratégies de réussite dans l'enseignement supérieur*, Dakar, 5-7 avril 2001.
- Detroz, P et Noël, F. (2002). Etude de l'impact de l'utilisation des boîtiers de votes en amphithéâtre, *actes du Colloque AIPU*. Louvain-La-Neuve: Université Catholique de Louvain.
- Gilles, J.-L. et Leclercq, D. (1995). Procédures d'évaluation adaptées à des grands groupes d'étudiants universitaires - Enjeux et solutions pratiquées à la FAPSE-ULG, in *Actes du Symposium International sur la Rénovation Didactique en Biologie*. Tunis : Université de Tunis.
- Gilles, J.-L., Bourguignon, J.-P. et Detroz, P., (2000). Les questionnaires à choix multiple : utilisation pour l'enseignement en groupe avec boîtiers électroniques. *Revue Médicale de Liège*, 55 : 12, pp. 1047-1050.
- Gilles, J.-L. et Melon, S. (2000). Comparaison de trois modalités de « testing » des compétences en français chez les étudiants médecins lors de leur première candidature à l'ULg, in Defays et al. (Eds), *La maîtrise du français du niveau secondaire au niveau supérieur*, Bruxelles : Editions De Boeck, pp. 161-178.
- Gilles, J.-L. (1997). Impact de deux entraînements à l'utilisation des degrés de certitude chez les étudiants de 1<sup>ère</sup> candidature de la Faculté de Psychologie et des Sciences de l'Education de l'ULg, in Boxus, E., Gilles, J.-L., Jans, V. & Leclercq, D. (Eds), *Actes du 15<sup>ème</sup> Colloque de l'Association Internationale de Pédagogie Universitaire (A.I.P.U.)*. Liège : Affaires Académiques de l'Université de Liège, pp. 311-326.
- Gilles, J.-L. (1998). Mise en œuvre de tests formatifs à l'aide de l'internet, in Depover & Noël (Eds), *Approches plurielles de l'évaluation des processus cognitifs*. Mons : Université de Mons-Hainaut, pp. 193-204.
- Gilles, J.-L. (2002). Qualité spectrale des tests standardisés universitaires – Mise au point d'indices éduométriques 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.
- Gilles, J.-L. (2004). Feedbacks diagnostiques via internet : un système individualisé de communication des niveaux de performances des étudiants dans le cadre d'évaluations standardisées, in *Actes du 3<sup>ème</sup> Congrès des chercheurs en éducation « (Re)trouver le plaisir d'enseigner et d'apprendre – Construire savoirs et compétences »*, Bruxelles : Administration générale de l'Enseignement et de la recherche scientifique (à paraître).
- Jans & Leclercq (1999). Mesurer l'effet de l'apprentissage à l'aide de l'analyse spectrale des performances, in C. Depover & B. Noel (Ed.), *Evaluation des compétences et des processus cognitifs*. Bruxelles : De Boeck, pp. 303-317
- Leclercq, D. (1993). Validity, Reliability, And Acuity of Self-Assessment in Educational Testing. In D. Leclercq et J. Bruno (Eds), NATO ASI Series, *Item Banking: Interactive Testing and Self Assessment*. Berlin: Springer Verlag, Vol. 112, pp. 114-131.