

PROJET DE PLATEFORME ELECTRONIQUE  
DE CONSTRUCTION ET DE GESTION QUALITE DES TESTS STANDARDISES  
*Programme WIST - Convention 031/5437 (liée à la convention 031/5585)*

## **e-C&QCST**

*electronic CONSTRUCTION & QUALITY CONTROL  
in STANDARDIZED TESTING*

# 3EME RAPPORT SCIENTIFIQUE DE L'EQUIPE DU SYSTEME METHODOLOGIQUE D'AIDE A LA REALISATION DE TESTS (SMART)

JEAN-LUC GILLES ET PASCAL DETROZ



Projet de recherche soutenu par la  
Direction Générale des Technologies, de la Recherche et de l'Energie (DGTRE)  
de la Région Wallonne

Décembre 2006



## Système Méthodologique d'Aide à la Réalisation de Tests

Responsable académique : Prof. Jean-Luc GILLES  
Directeur : Pascal DETROZ



## SOMMAIRE

### Préambule :

1. Vers une pérennisation du projet : les enjeux liés à l'extension du travail R&D en 2007	5
2. Identification du projet	10
3. Identification du SMART-ULg	12
<b>I. INTRODUCTION</b>	<b>16</b>
1. <b>Rappel des bases scientifiques et des objectifs qui sous-tendent le projet <i>e-C&amp;QCST</i></b>	<b>16</b>
a) Les bases scientifiques sur lesquelles repose le lancement du projet	16
b) Les objectifs de l'ensemble du projet	16
2. <b>Rappel de l'état des connaissances et publications au niveau national et international dans le domaine du projet</b>	<b>17</b>
3. <b>Rappel de l'état des connaissances des partenaires dans le domaine du projet</b>	<b>19</b>
4. <b>Rappel de l'innovation technologique apportée par la plate-forme <i>e-C&amp;QCST</i></b>	<b>20</b>
<b>II. RAPPEL DE NOTRE VISION SCIENTIFIQUE DE DEPART EN CE QUI CONCERNE L'APPROCHE QUALITE EN TESTING STANDARDISE</b>	<b>23</b>
1. <b>Pénétration des concepts « Qualité » dans les activités de testing standardisé du Système Méthodologique d'Aide à la Réalisation de Tests (SMART) de l'Université de Liège</b>	<b>23</b>
a) Le contexte du SMART	23
b) Historique du SMART	24
c) Facteurs de mutation de l'université depuis les années 60	26
b) Définir le concept de « Qualité »	28
2. <b>Critères de qualité des évaluations standardisées</b>	<b>28</b>
a) La validité	29
b) La fidélité	37
c) La sensibilité	39
d) La diagnosticité	40
e) La praticabilité	41

f)	L'équité	41
g)	La communicabilité	42
h)	L'authenticité	42
<b>3.</b>	<b>Cycle de construction en « spirale de qualité » des évaluations standardisées et dispositifs d'ingénierie docimologique associés</b>	<b>43</b>
a)	Analyse des objectifs de l'enseignement ou de la formation	44
b)	Mise en forme de l'épreuve	46
c)	Construction de l'épreuve	47
d)	Entraînement des évalués	48
e)	Mise en œuvre de l'épreuve	49
f)	Correction et discussion des résultats	52
g)	Feedbacks aux évalués après l'épreuve	55
h)	Macro régulation	57
<b>III.</b>	<b>PRESENTATION DE L'ÉTAT DE LA RECHERCHE E-C&amp;QCST A LA 11EME CONFERENCE BIENNALE DE L'EUROPEAN ASSOCIATION FOR RESEARCH IN LEARNING AND INSTRUCTION (EARLI)</b>	<b>64</b>
<b>1.</b>	<b>electronic Construction and Quality Control in Standardized Testing, the e-C&amp;QCST project</b>	<b>65</b>
a)	Introduction	65
b)	Method	67
c)	First results	70
d)	Conclusions	76
e)	References	78
<b>2.</b>	<b>Development of edumetrical indices for the analysis of the spectral quality of higher education standardized tests</b>	<b>79</b>
a)	Introduction	80
b)	Method	82
c)	Results	87
d)	Conclusions	92
e)	References	95
<b>IV.</b>	<b>e-C&amp;QCST COMMUNITY : LES PREMISES D'UNE COMMUNAUTE D'UTILISATEURS DE LA PLATEFORME DE CONSTRUCTION ET DE GESTION QUALITE DES TESTS STANDARDISES</b>	<b>96</b>
<b>1.</b>	<b>Pertinence du concept de "Communauté d'utilisateurs"</b>	<b>96</b>
a)	Eviter de perpétuer avec un nouvel outil d'anciennes pratiques	96
b)	Favoriser une démarche de prise en main innovante	96

<b>2. Premières interfaces</b>	<b>97</b>
a) Canevas de la page principale	97
b) Espace réservé à l'équipe	98
c) Espace d'information publique	98
<b>3. Premières utilisations</b>	<b>99</b>
<b>V. ASPECTS SCIENTIFIQUES LIES A LA VALORISATION DE LA RECHERCHE e-C&amp;QCST</b>	<b>100</b>
<b>1. Contexte de la valorisation économique des services de testing standardisés</b>	<b>100</b>
a) Contexte socio-pédagogique	100
b) Contexte socio-politique	100
c) Contexte scientifique	101
d) Contexte économique	102
<b>2. Le projet First Spin Off « Smart Advices For Evaluation Systems (SAFES) »</b>	<b>103</b>
a) Problématique de départ en lien avec e-C&QCST	103
b) Description du projet SAFES	104
c) Objectifs poursuivis et approche méthodologique	105
d) Innovation technologique apportée par le projet SAFES	111
e) Calendrier de réalisation	113
<b>3. De la recherche à la consultance : le choix stratégique d'un transfert d'activité opéré par l'équipe du Système Méthodologique d'Aide à la Réalisation de Tests (SMART) de l'Université de Liège</b>	<b>114</b>
a) Introduction	114
b) L'afflux des demandes de prestations pour tiers	116
c) Gérer la valorisation des R&D dans le contexte du SMART	118
d) Le scenario de la spin off « Assess Group Smart Advices For Evaluation Systems » (SAFES)	123
e) Conclusions	125
<b>VI. OUTPUT DU PROJET : LA PLATEFORME EXAMS</b>	<b>128</b>
<b>Quick Start Guide</b>	<b>129</b>

<b>VII. CONCLUSIONS</b>	<b>167</b>
<b>1. Effets en retour du développement informatique sur nos conceptions docimologiques</b>	<b>167</b>
a) Clarification et affinement des concepts liés au cycle CGQTS	<b>167</b>
b) La nécessaire formalisation et gestion informatisée des écarts entre tâches docimologiques prescrites et tâches docimologiques effectuées	<b>168</b>
c) De nouvelles possibilités qui stimulent la pensée créative du chercheur en docimologie	<b>168</b>
<b>2. Perspectives futures</b>	<b>169</b>
a) Adapter la plate-forme exAMS à l'évaluation et à la gestion des compétences	<b>169</b>
b) La nécessaire formalisation et gestion informatisée des écarts entre tâches docimologiques prescrites et tâches docimologiques effectuées	<b>171</b>
<b>3. Conclusion générale</b>	<b>173</b>
<b>VIII. BIBLIOGRAPHIE</b>	<b>175</b>
<b>IX. ANNEXES</b>	<b>185</b>
Annexe I : Dias PowerPoint de la présentation « electronic Construction and Quality Control in Standardized Testing, the <i>e-C&amp;QCST</i> project » effectuée à la 11 <sup>ème</sup> conférence biennale de EARLI	<b>186</b>
Annexe II : Dias PowerPoint de la présentation « Development of edumetrical indices for the analysis of the spectral quality of higher education standardized tests » effectuée à la 11 <sup>ème</sup> conférence biennale de EARLI	<b>199</b>
Annexe III : Convention de recherche First Spin Off « SAFES »	<b>215</b>

## Préambule

### 1. Vers une pérennisation du projet : les enjeux liés à l'extension du travail R&D en 2007

Fin décembre 2006, à l'aide de fonds propres, l'équipe du Système Méthodologique d'Aide à la Réalisation de Tests (SMART) a pris la décision de poursuivre les travaux de R&D relatifs à la recherche *electronic Construction and Quality Control in Standardized Testing (e-C&QCST)*. Ce choix est dicté par une série de raisons que nous allons détailler dans les paragraphes qui suivent. Signalons ici que l'énergie consacrée à *e-C&QCST* en 2007 équivaut à 2 ETP programmeurs et 0,5 ETP chercheurs docimologues.

#### *1.1 Evolution des paradigmes d'ingénierie docimologiques*

Une des raisons de la prolongation du travail sur l'année 2007 réside dans l'évolution des conceptions docimologiques en matière de construction et de gestion de la qualité des épreuves suite aux interactions des docimologues avec les partenaires de terrain et les informaticiens de l'équipe.

Le travail avec les partenaires de terrain a mis en évidence des besoins en termes de fonctionnalités docimologiques qui n'étaient pas apparus au début du projet. Ceci s'explique par le fait que l'utilisation de la plate-forme par nos partenaires (en particulier AXA, FORTIS BANK, et IFA) à fait naître chez eux des demandes qui n'auraient jamais émergées sans une utilisation par leurs services des premières versions de la plate-forme. L'exemple qui illustre le mieux ce phénomène d'évolution des besoins en interaction avec l'usage de la plate-forme est celui des fonctionnalités de l'étape IV « *Training* ». Dans la version initiale, il s'agissait essentiellement de proposer des entraînements à la passation de tests. Après quelques utilisations dans la phase d'expérimentation, il est apparu qu'il fallait, au-delà de ces entraînements, proposer bien d'autres fonctionnalités de mise en ligne d'informations relatives à l'épreuve : des explications détaillées des consignes, des informations sur les conditions de passation de l'épreuve, de la documentation sur la matière qui serait concernée par l'évaluation (notamment des possibilités de visualisation d'une partie de la table de spécification créée à l'étape 1 « *Analyse* »), ... Bref, toutes sortes de documents permettant de mieux rencontrer notre objectif qualité « *Communicabilité* » selon lequel les informations non confidentielles relatives au déroulement du processus d'évaluation doivent être communiquées et comprises par les personnes engagées dans la réalisation de ces évaluations (voir infra).

Le même type de phénomène d'émergence de besoins nouveaux suite aux interactions avec des partenaires du projet s'est présenté dans le contexte du dialogue collaboratif avec l'équipe des informaticiens. En programmant les interfaces proposées par les docimologues, les informaticiens ont été amenés à envisager des cas limites dans l'usage de la plate-forme. La réflexion engendrée par l'analyse de ces cas limites à fait ressortir de nouvelles possibilités docimologiques qui à l'origine du projet n'avaient pas été entrevues.

Par exemple, dans la gestion des feedbacks aux évalués après une épreuve, différents moments ont été formalisés et croisés avec différents types de feedbacks, ce qui donne le tableau qui suit où, selon le type d'épreuve (électronique en ligne ou « papier crayon »), certains cas de

figure sont envisageables (cases à cocher) ou non (croix marquant les impossibilités). Les choix finaux pour les cases qui peuvent être cochées et qui désignent les possibilités offertes étant du ressort de l'évaluateur lorsqu'il conçoit son épreuve à l'étape 2 « *Design* » du processus de construction.

Choix des Feedbacks :	Moment 1 : Immédiatement après la question	Moment 2 : Immédiatement à la fin du test	Moment 3 : Dès que le test a été passé par tous les évalués	Moment 4 : Après discussion sur base de la version définitive
FB1 sur l'exactitude des solutions proposées	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
FB2 expliquant en quoi les solutions proposées sont correctes ou non	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
FB3 partiel sur l'état des performances globales de l'évalué question après question	<input checked="" type="checkbox"/>	-	-	-
FB4 sur les performances globales de l'évalué pour l'ensemble ou les différentes parties du test	-	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
FB5 des performances du groupe pour l'ensemble ou les différentes parties du test	-	-	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
FB6 sur la qualité des résultats aux propositions, questions et test	-	-	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>

Evidemment, suite à notre décision de prolonger le travail de développement, les interactions entre l'équipe des docimologues, les partenaires de terrain et l'équipe informatique continuent et font naître de nouvelles fonctionnalités qui elles-mêmes font émerger de nouveaux besoins. Autrement dit, lors des derniers mois de l'année 2006, lorsque la plate-forme a pu commencer à exprimer toute la puissance docimologique de ses fonctionnalités et que les partenaires ont pu en tirer des bénéfices d'usage sur leur terrains d'expérimentations, l'équipe du projet a été entraînée dans une très riche spirale d'élaborations conceptuelles de nouvelles fonctionnalités d'ingénierie docimologique résultant de ces interactions partenaires-docimologues-informaticiens. Du point de vue de la recherche docimologique il convenait de ne pas interrompre cette spirale et de mettre tout en œuvre pour en formaliser les résultats dans la plate-forme car ce qui se dessinait progressivement, c'est l'émergence d'un véritable Assessment Management System (AMS) dont nous reparlerons plus loin.

### ***1.2 Intégration de la plate-forme dans les services docimologiques proposés par le SMART au sein de l'ULg***

Une seconde raison expliquant la nécessité de prolonger sur fonds propres le développement de la plate-forme issue de cette recherche est liée à la mission principale de l'équipe du SMART, à savoir l'accompagnement méthodologique et logistique des enseignants de l'Université de Liège dans la réalisation de leurs évaluations de grands groupes d'étudiants.

La mise en œuvre de la plate-forme de construction et de gestion de la qualité des tests standardisés dans le contexte de l'ULg nous a amené à devoir créer les conditions de son interopérabilité avec d'autres logiciels développés au sein du SMART pour assurer les aspects logistiques de l'aide aux enseignants. Par exemple, pour satisfaire les besoins de dizaines de professeurs au sein de l'institution, le SMART utilise un lecteur optique de marques (LOM) pour l'acquisition automatique des réponses de grands groupes d'étudiants. Ce dispositif LOM permet de fournir en peu de temps des fichiers de données de réponses de centaines d'étudiants. Ces données sont ensuite traitées par des logiciels développés depuis plusieurs années par le SMART en vue de fournir rapidement les résultats des examens aux enseignants. Le logiciel qui a été créé à la fin des années '90 pour effectuer ces traitements et permettre aux enseignants via internet porte le nom « Traitement Via Internet » (TVI). D'un côté, ce logiciel permet également de contrôler la qualité des questions d'un test mais avec beaucoup moins de possibilités et de souplesse que la nouvelle plate-forme issue de la recherche e-C&QCST. Par ailleurs, TVI est un logiciel qui n'intervient qu'aux étapes « 6. Correction » et « 7. Feedbacks » et qui ne couvre pas les autres étapes du cycle de Construction et de Gestion Qualité des Tests Standardisés (CGQTS) instrumenté par la nouvelle plate-forme. D'un autre côté, TVI est conçu pour gérer les données issues de la lecture optique en tenant compte des spécifications du matériel utilisés par le SMART (un lecteur AXM 990), ce qui n'est pas envisagé par la nouvelle plate-forme qui n'est pas conçue pour gérer un type de matériel particulier.

Dès lors, l'enjeu pour l'équipe de développement était d'une part de rendre la nouvelle plate-forme compatible avec l'ancien logiciel TVI dont il fallait modifier certains aspects pour permettre une communication efficace. D'autre part, il s'agissait d'intégrer dans le contexte de cette démarche de mise en compatibilité des deux logiciels, une réflexion plus profonde sur les conditions d'une interopérabilité large avec d'autres systèmes logistiques d'acquisition automatique de réponses à l'étape « 6. Correction » et, partant des solutions mises en œuvre pour cette étape, d'étendre les capacités d'interopérabilité du système aux sept autres étapes du cycle CGQTS.

Ce travail débouche aujourd'hui sur des solutions opérationnelles qui viennent d'être mises en œuvre dans le contexte institutionnel du SMART et de l'évaluation des grands groupes d'étudiants au sein de l'Université de Liège. La mise en œuvre de ces solutions a par ailleurs permis de faire évoluer la plate-forme vers une plus grande interopérabilité avec les multiples solutions logicielles utilisées dans les entreprises et les grandes administrations, augmentant ainsi considérablement les possibilités d'utilisation dans des contextes différents.

### ***1.3 Partenariats avec la spin off Assess-Group S.A. dans le cadre d'une offre de services docimologiques pour de grands comptes***

L'équipe du SMART a fait le choix d'externaliser ses activités de prestations pour tiers en créant en octobre 2006 la spin off Assess-Group S.A. Ce choix stratégique est expliqué en détail dans la partie « V. ASPECTS SCIENTIFIQUES LIÉS À LA VALORISATION DE LA RECHERCHE e-C&QCST » du présent rapport. Aujourd'hui, cette spin off occupe déjà 6 personnes dans le secteur d'activité lié à l'évaluation et compte parmi ses clients de grands comptes tels que FORTIS BANK, BELGACOM, AXA, ... ainsi que de grandes administrations telles que l'IFA, le FOREM, des ministères de la République Française, ... Récemment, des contacts ont été pris avec la Chine dans le cadre de la mission économique de la Province de Liège, deux membres de la spin off (dont un ex-membre de l'équipe e-C&QCST) faisaient partie de la délégation qui s'est

rendue au Fujian en juin 2007.

Assess Group S.A. valorise les résultats des recherches du SMART dans ses trois missions et donc aussi dans le cadre de l'évaluation standardisée des acquis. Dans le contexte de cette mission, les offres de services déployées dans le domaine de la construction et de la gestion de la qualité des tests standardisés gagnent à pouvoir s'appuyer sur les fonctionnalités offertes par la plate-forme « open source » issue de la recherche e-C&QCST. Dès lors, nous avons institué un dialogue collaboratif étroit entre l'équipe SMART de développement de la plate-forme et l'équipe Assess-Group S.A. de valorisation des services d'accompagnements méthodologiques, techniques et logistiques liés à son utilisation par des tiers.

Dès la fin 2006, une première collaboration liée à la mise en œuvre de la plate-forme a concrètement vu le jour entre le SMART et la spin off pour répondre à une demande de services émanant de la firme AXA. Cette banque ayant émis le souhait d'être accompagnée par Assess-Group S.A. dans le cadre de la mise en œuvre de nouveaux processus de gestion et d'évaluation des compétences de son personnel. Dans ce contexte, la spin off a proposé l'utilisation de la plate-forme en vue d'instrumenter les procédures de gestion des épreuves envisagées. Connaissant déjà la plate-forme (étant donné le partenariat instauré dans le cadre de la recherche e-C&QCST) AXA a rapidement accepté l'idée et un contrat de services docimologiques instrumentés à l'aide de la plate-forme fut signé peu de temps après entre la spin off et AXA.

Dès les premières réunions de travail avec AXA, il s'est avéré nécessaire de modifier certaines fonctionnalités de la plate-forme afin de permettre une meilleure adaptation aux contraintes du nouveau projet de la banque. Cette demande d'adaptation de la part d'AXA vis-à-vis d'Assess Group S.A. entraîna d'autres de la part de la spin off vis-à-vis de l'équipe de développement du SMART et une nouvelle réflexion en profondeur fut entamée pour faciliter l'intégration de la plate-forme dans le contexte de grands comptes tels qu'AXA. L'idée étant de répondre aux besoins d'AXA tout en proposant des solutions qui pourraient être étendues à d'autres contextes de grandes entreprises ou administrations.

A nouveau, ces études liées aux réponses à fournir dans le cas d'AXA, mais dans une perspective de généralisation à d'autres contextes similaires, ont entraîné des refontes en profondeur du code de la plate-forme. Cela ne pouvait se faire qu'en prolongeant la R&D sur l'année 2007, ce que nous avons fait dans le cas d'AXA en finançant une partie du travail à l'aide d'une enveloppe budgétaire octroyée par AXA à la spin off qui a ensuite sous-traité le développement à l'équipe du SMART.

Aujourd'hui la plate-forme en ressort consolidée, non seulement par le fait qu'une grande banque telle qu'AXA l'utilise déjà en « grandeur réelle » mais aussi parce que le code de la plate-forme est plus solide et que ses fonctionnalités sont encore plus adaptées aux réalités du terrain des grands comptes tout en offrant la souplesse et la flexibilité nécessaire pour répondre à d'autres besoins émanant de petites et moyennes entreprises.

### ***1.4 La mise aux normes pour un travail collaboratif dans une communauté de développeurs***

Notre soucis pendant cette année 2007, fut aussi de préparer le mieux possible l'avenir de la plate-forme sachant que les possibilités de développements futurs passeront par un travail en

collaboration étroite avec des développeurs qui viendront ponctuellement travailler sur le projet. Ceux-ci le feront le plus souvent pour répondre à des besoins qui leurs sont propres (souvent liés à l'interopérabilité avec d'autres logiciels en place dans les entreprises et les institutions) mais à l'aide d'un système que nous mettons en place et qui permettra une intégration cohérente avec ce qui existe déjà. Ce système doit également permettre le partage et la redistribution des nouvelles fonctionnalités créées au sein d'une communauté d'utilisateurs. Dans ce scénario, le SMART reste la structure mère centralisatrice des développements futurs et garante auprès de la communauté de la qualité et de la pertinence de ceux-ci.

La mise en place de cette vision partagée des développements futurs nécessitait qu'on teste et qu'on formalise nos hypothèses de travail lors d'une expérience pilote. Cela s'est concrétisé au travers du travail de fin d'étude d'un étudiant en informatique que nous avons placé dans la situation d'un développeur externe à l'équipe e-C&QCST et à qui était confié une tâche de programmation précise. Cette tâche consistait en la réalisation d'une interface permettant d'utiliser les tables de spécifications créées au sein de la plate-forme dans un autre environnement : le Learning Management System (LMS) DOKEOS qui est également « open source ».

Ce nouveau développement pris en charge par une personne extérieure à l'équipe et qui n'avait jusque là jamais participé à notre projet, a permis, au-delà de la « simple » réalisation des nouvelles fonctionnalités demandées, de restructurer le code de la plate-forme en profondeur de façon à améliorer l'intégration des apports de collaborateurs externes ponctuels faisant partie d'une communauté de développeurs.

Notons, qu'au-delà de cet apport essentiel à la mise en place d'une large communauté de développeurs, le travail effectué par l'étudiant encadré par l'équipe e-C&QCST a permis de rapprocher deux mondes : celui des Learning Management Systems (LMS) et celui des Assessment Management Systems (AMS), ce qui à notre connaissance constitue une première mondiale dans le contexte des applications « open source ».

## 2. Identification du projet

**1.1 Titre du Projet :** plate-forme électronique de construction et de gestion qualité de tests standardisés (*electronic Construction & Quality Control in Standardized Testing*)  
**Acronyme:** e-C&QCST

**1.2 Contrat/Convention** 031/5437 (liée à la convention 031/5585)  
**Durée de la recherche (mois) : 38 mois**

**1.3 Date de commencement du Projet**  
01/11/03 en ce qui concerne l'ULg (le projet a débuté le 01/01/04 aux HEC)

**1.4 Période couverte par le rapport d'activité**  
01/11/04 au 31/12/05

### 1.5 Coordinateur et/ou Chef de Projet

**Prof. Dr. Jean-Luc Gilles**

Responsable académique de l'Unité de Soutien Logistique et de Recherche du Système Méthodologique d'Aide à la Réalisation de Tests (USLR-SMART)

Institut de Formation et de Recherche en Enseignement Supérieur (IFRES)

Université de Liège

Adresse : Traverse des architectes, 4 (Bât. B3c)

CP : 4000

Localité : Liège (Sart Tilman)

Tél. : 04-366.20.78

Fax : 04-366.27.00

E-mail : jl.gilles@ulg.ac.be

Chef de service de l'Unité de Didactique Générale et Intervention Educative (DGIE)

Faculté de Psychologie et des Sciences de l'Education (FAPSE)

Université de Liège

Adresse : Boulevard du Rectorat, 4 (Bât. B32)

CP : 4000

Localité : Liège (Sart Tilman)

Tél. : 04-366.45.88

E-mail : jl.gilles@ulg.ac.be

### 1.6 Liste des différents Partenaires

Partenaire 1 (coordinateur, chef de projet) – Gilles Jean-Luc (ULg, USLR-SMART)

Partenaire 2 – Dujardin Jean-Marie (HEC Liège, cellule Compétences)

### 1.7. Rôle de chacun des partenaires

(1) *Le système méthodologique d'aide à la réalisation de tests de l'Université de Liège (SMART-ULg)*

Le Système Méthodologique d'Aide à la Réalisation de Tests (SMART) est une équipe pluridisciplinaire spécialisée dans trois domaines liés à l'évaluation pédagogique :

- L'évaluation des acquis des apprenants : Aider les enseignants et les formateurs à réaliser les évaluations de leurs apprenants (exercices, examens, tests formatifs...), spécialement pour les grands groupes.
- Les audits qualité en formation : Aider les organismes de formation à évaluer la qualité de leurs formations, à mener de grandes enquêtes de satisfaction auprès des formés.
- Les enquêtes et surveys pédagogiques : Aider les institutions d'enseignement, les administrations ou les entreprises à réaliser des opérations de collecte de données, plus particulièrement lorsqu'elles ont un caractère pédagogique.

Le SMART, cellule spécialisée en évaluation standardisée de l'Université de Liège, est responsable de cette recherche et s'est associé avec le département Gestion des Compétences des HEC pour mener cette recherche-action.

Le SMART apporte son expertise en docimologie et en évaluation standardisée ainsi qu'un ensemble de procédures mises à l'épreuve à l'université, dans diverses hautes écoles et dans des entreprises de formation. Il est le dépositaire du modèle théorique sous-tendant l'ensemble de cette plateforme d'évaluation (Gilles, 2002) . Il coordonne la recherche et a réaliser la partie informatique du projet.

La cellule de gestion des compétences des HEC - ULg est le partenaire du projet

Description :

### *(2) HEC- cellule de gestion des compétences*

La cellule Gestion des Compétences de l'UER Management développe des activités de recherches dans le domaine de la gestion des compétences, l'évolution des compétences sur le marché du travail, la gestion de la formation en entreprise, l'évaluation et la validation de compétences.

Ces projets sont commandités pour la plupart par des pouvoirs publics (Région wallonne et Union Européenne). Nous privilégions les recherches-actions, associant des problématiques concrètes de terrain à une analyse de la littérature et à des rencontres d'experts et aboutissant, in fine, à des productions utilisables (outils de gestion de ressources humaines, logiciel, etc.) et/ou à une diffusion la plus large possible des résultats (conférences, articles, colloques, etc.).

L'équipe de recherche est multi-disciplinaire : management, psychologie du travail et des organisations, et pédagogie. Nos projets sont menés en partenariat avec des institutions universitaires, des entreprises privées ou des organisations non-marchandes, choisies pour leurs apports complémentaires ou leurs intérêts de recherche communs.

Rôle dans le projet :

La cellule de recherche « Gestion des Compétences » de HEC - Ecole de Gestion de l'Université de Liège a eu pour mission de proposer d'éventuelles adaptations des démarches proposées et des caractéristiques de la plate-forme par rapport à une utilisation de l'outil en entreprise.

## 2. Identification du SMART – ULg

### *Système Méthodologique d'Aide à la Réalisation de Tests*

#### **2.1 Coordonnées du partenaire SMART - ULg :**

**Prof. Dr. Jean-Luc Gilles**

SMART

Université de Liège

Adresse : Traverse des Architectes, 4 (Bât. B3c)

CP : 4000

Localité : Liège (Sart Tilman)

Tél. : 04-366.22.80

Fax : 04-366.27.00

E-mail : jl.gilles@ulg.ac.be

#### **2.2 Chercheurs et Techniciens engagés par le partenaire SMART-ULg :**

##### ***Chercheur 1***

**Prénom Nom : Jean-Luc Gilles**

Qualification : docteur en Sciences de l'Éducation, spécialisé en docimologie

Fonction dans cette recherche : chef de projet et coordinateur pour l'ULg

Date d'engagement (pour cette recherche): 01/11/03

Tél. : 04.366.22.80

Fax : 04.366.27.00

E-mail : jl.gilles@ulg.ac.be

##### ***Chercheur 2***

**Prénom Nom : Pascal Detroz**

Qualification : licencié en Psychologie, maîtrise en ergonomie

Fonction dans cette recherche : chercheur

Date d'engagement (pour cette recherche): 01/11/03

Tél. : 04.366.22.81

Fax : 04.366.27.00

E-mail : p.detroz@ulg.ac.be

##### ***Chercheur 3***

**Prénom Nom : Myriam Dupont**

Qualification : licenciée en Psychologie

Fonction dans cette recherche : chercheuse

Date d'engagement (pour cette recherche): 01/01/06

Tél. : 04.366.22.80

Fax : 04.366.27.00

E-mail : myriam.dupont@ulg.ac.be

***Chercheur 4***

***Prénom Nom : Audrey Mohr***

Qualification : licenciée en Sciences de l'Éducation

Fonction dans cette recherche : chercheuse

Date d'engagement (pour cette recherche): 01/01/06

Tél .: 04.366.22.80

Fax : 04.366.27.00

E-mail : amohr@ulg.ac.be

***Chercheur 5***

***Prénom Nom : Vinciane Crahay***

Qualification : graduée

Fonction dans cette recherche : chercheuse

Date d'engagement (pour cette recherche): 01/09/06

Tél .: 04.366.20.78

Fax : 04.366.27.00

E-mail : v.crahay@ulg.ac.be

***Technicien 1***

***Prénom Nom : Marc Pirson***

Qualification : gradué en Informatique

Fonction dans cette recherche : analyste-programmeur spécialisé dans le langage PHP, les bases de données MySQL et la gestion de serveur Apache

Date d'engagement (pour cette recherche): 01/11/03

Tél .: 04.366.22.82

Fax : 04.366.27.00

E-mail : [m.pirson@ulg.ac.be](mailto:m.pirson@ulg.ac.be)

***Technicien 2***

***Prénom Nom : Mbassa Dabo***

Qualification : licencié en Informatique

Fonction dans cette recherche : informaticien spécialisé dans le langage PHP et les bases de données MySQL

Date d'engagement (pour cette recherche): 01/06/04 jusqu'au 30/05/06

Tél .: 04.366.22.82

Fax : 04.366.27.00

E-mail : mdabo@ulg.ac.be

***Technicien 3***

***Prénom Nom : Hung LÊ***

Qualification : gradué en Informatique

Fonction dans cette recherche : analyste-programmeur spécialisé dans le langage PHP et les bases de données MySQL

Date d'engagement (pour cette recherche): 08/11/04

Tél .: 04.366.22.82

Fax : 04.366.27.00

E-mail : PH.Le@ulg.ac.be

***Technicien 4***

***Prénom Nom : Jean-Philippe Humblet***

Qualification : gradué en Informatique

Fonction dans cette recherche : analyste-programmeur spécialisé dans le langage PHP et les bases de données MySQL

Date d'engagement (pour cette recherche): 01/01/06

Tél .: 04.366.22.82

Fax : 04.366.27.00

E-mail : jp.humblet@ulg.ac.be

***Technicien 5***

***Prénom Nom : Yannick Lescalier***

Qualification : gradué en Informatique

Fonction dans cette recherche : analyste-programmeur spécialisé dans le langage PHP et les bases de données MySQL

Date d'engagement (pour cette recherche): 01/01/06

Tél .: 04.366.22.82

Fax : 04.366.27.00

E-mail : yannick.lescalier@ulg.ac.be

***Technicien 6***

***Prénom Nom : Frédéric GILLIQUET***

Qualification : gradué en Informatique

Fonction dans cette recherche : analyste-programmeur spécialisé dans le langage PHP et les bases de données MySQL

Date d'engagement (pour cette recherche): 05/04/06

Tél .: 04.366.22.82

Fax : 04.366.27.00

E-mail : [fr.gilliquet@ulg.ac.be](mailto:fr.gilliquet@ulg.ac.be)

***Technicien 7***

***Prénom Nom : Stephane Orban***

Qualification : gradué en Informatique

Fonction dans cette recherche : analyste-programmeur spécialisé dans le langage PHP et les bases de données MySQL

Date d'engagement (pour cette recherche): 01/01/06

Tél . : 04.366.22.82

Fax : 04.366.27.00

E-mail : [s.orban@ulg.ac.be](mailto:s.orban@ulg.ac.be)

***Technicien 8***

***Prénom Nom : Damien Delpuvrez***

Qualification : gradué en Informatique

Fonction dans cette recherche : analyste-programmeur spécialisé dans le langage PHP et les bases de données MySQL

Date d'engagement (pour cette recherche): 01/07/06

Tél . : 04.366.22.82

Fax : 04.366.27.00

E-mail : [damien.depluvrez@ulg.ac.be](mailto:damien.depluvrez@ulg.ac.be)

***Technicien 9***

***Prénom Nom : Frederic Bruyère***

Qualification : gradué en Informatique

Fonction dans cette recherche : analyste-programmeur spécialisé dans le langage PHP et les bases de données MySQL

Date d'engagement (pour cette recherche): 01/08/06

Tél . : 04.366.22.82

Fax : 04.366.27.00

E-mail : [damien.depluvrez@ulg.ac.be](mailto:damien.depluvrez@ulg.ac.be)

## I. INTRODUCTION

*Prof. Dr. Jean-Luc GILLES,  
Coordinateur, chef de projet e-C&QCST*

### 1. Rappel des bases scientifiques et des objectifs qui sous-tendent le projet e-C&QCST

#### a) Les bases scientifiques sur lesquelles repose le lancement du projet

Le testing standardisé est une forme d'évaluation des compétences des apprenants qui, lorsqu'elle réunit les qualités de validité, fidélité, sensibilité, diagnosticité, équité, praticabilité et communicabilité (Gilles & Leclercq, 1997 ; Castaigne, Gilles & Hansen, 2001 ; Gilles, 2002), est appréciée à la fois par les formateurs, les formés et les instances dirigeantes des institutions ou organismes de formation.

Les préoccupations relatives à la qualité des évaluations standardisées des acquis des formés s'insère dans un mouvement qui depuis le début des années '90, touche les institutions belges et européennes (Kaufman & al., 1995 ; Segers & Dochy, 1996). Ce mouvement participe lui-même à une tendance mondiale d'introduction du management de la qualité dans la sphère des activités d'enseignement et de formation (Ramsden, 1991 ; Nightingale & O'Neil, 1994 ; Zink & Schmidt, 1995).

Conscients de ces lignes de force actuelles, nous proposons une recherche visant la mise en place de méthodes et d'outils de gestion de la qualité des épreuves standardisées en vue de satisfaire au mieux les exigences des enseignants/formateurs et des étudiants/formés européens, voire d'autres continents. La dissémination et l'utilisation à l'échelle mondiale des résultats de notre recherche sont aujourd'hui possibles à l'aide d'interfaces conçues pour fonctionner via le web. C'est l'idée de notre projet e-C&QCST<sup>1</sup> de plate-forme électronique de construction et de gestion qualité de tests standardisés réalisée à l'aide de logiciels libres et dans une philosophie « *Open Source* ».

#### b) Les objectifs de l'ensemble du projet

Il s'agit d'abord de créer des procédures docimologiques qui permettent de maîtriser, quantifier et contrôler l'atteinte des objectifs qualité dans le cadre des étapes de construction d'une évaluation standardisée des compétences d'apprenants. Ensuite, ces procédures seront programmées. Les interfaces de la plate-forme e-C&QCST devront être accessibles via l'Internet et seront réalisées à l'aide de logiciels libres tels que le langage de programmation PHP, le gestionnaire de bases de données MySQL et un serveur web Apache.

---

<sup>1</sup> *electronic Construction & Quality Control in Standardized Testing*

L'architecture de notre plate-forme électronique de construction et de gestion qualité de tests standardisés repose sur huit modules intégrés :

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

Tout comme le recours à des méthodes et des dispositifs de testing utilisant les Technologies de l'Information et de la Communication (TIC) est incontournable pour répondre efficacement aux problèmes d'évaluation standardisée; la mise en œuvre d'une gestion qualité de ces tests s'appuyant sur les TIC est essentielle pour garantir un service optimal adapté aux exigences des enseignants/formateurs et des étudiants/formés, et, au-delà, des organismes de formation privés ou publics.

## 2. Rappel de l'état des connaissances et publications au niveau national et international dans le domaine du projet

Nous assistons depuis une décennie à un mouvement mondial d'introduction du management de la qualité dans la sphère des activités de l'enseignement supérieur (Ramsden, 1991 ; Nightingale & O'Neil, 1994 ; Zink & Schmidt, 1995). Les pratiques évaluatives dans les institutions d'enseignement supérieur n'échappent pas à cette tendance récente, c'est ce que soulignent Blais & al. (1997) : « *Dans le contexte social actuel les établissements d'enseignement supérieur sont de plus en plus souvent amenés à faire la démonstration de la qualité de la formation dispensée. En ce sens, l'évaluation des apprentissages, comme mécanisme de régulation et de contrôle exige une attention particulière. Par ailleurs, la régulation et le contrôle des apprentissages doivent prendre en considération un souci de plus en plus net d'assurer l'équité et la transparence des pratiques d'évaluation* ». Notre recherche se situe dans cette réflexion sur la mise en place de méthodes de gestion de la qualité des évaluations permettant aux institutions de faire la démonstration de la qualité de leurs épreuves certificatives. Plus particulièrement, nous nous attachons à la qualité des tests standardisés de plus en plus utilisés comme « *mécanisme de régulation et de contrôle* » dans des institutions qui doivent faire face à divers de facteurs de mutation depuis les années '60 dont l'augmentation constante du nombre de personnes à former et donc à évaluer et certifier.

Parmi les divers facteurs de mutation de l'enseignement supérieur, Dupont & Ossandon (1994, p.13) citent notamment la diversification de l'origine sociale des étudiants : « ... *alors que les universitaires d'antan provenaient de filières secondaires sélectionnées, on observe que l'origine scolaire secondaire appartient à un éventail extrêmement varié de formations* ». A cette diversification sociale on peut aussi ajouter une tendance nouvelle de diversification des âges de ceux qui viennent se former à l'université. En effet, l'évolution rapide des sciences et des techniques entraîne la nécessité d'une formation continue.

On remarque également la création de passerelles de plus en plus nombreuses entre les formations, favorisant ainsi encore plus de diversité dans les publics. Ces nouveaux publics posent aussi de nouveaux défis en matière d'équité dans la certification des compétences. Dès lors, rappelons que si dans un pays comme les Etats-Unis les tests standardisés connaissent un tel succès, c'est notamment parce qu'ils offrent des garanties de non discrimination raciale (*culture-free test*) et de non discrimination de classe sociale (*class-free test*).

Et à ce constat de la diversification des publics, s'ajoute celui de la massification. Récemment à l'Université de Liège, des amphithéâtres pouvant contenir plus de 600 étudiants ont été construits. Dès lors, la pédagogie des grands groupes s'impose par elle-même et comme le soulignent Dupont & Ossandon (1994, p.15) : « *par la force des choses, la pédagogie universitaire, prise sous l'angle des méthodologies utilisées, est de plus en plus inadéquate ; la plupart des universités européennes sont confrontées à une pédagogie des grands groupes à laquelle la majorité des enseignants n'ont pas été sensibilisés ; on est loin d'un taux d'encadrement de 1 enseignant pour 15 à 20 élèves rencontré dans les prestigieuses universités américaines* ». Face à cette massification des effectifs, le testing standardisé s'impose comme une solution pragmatique pour la certification des compétences lorsque le contexte s'y prête. Par ailleurs, la pédagogie des grands groupes est souvent connotée négativement ce qui risque de déteindre sur les évaluations des standardisées des acquis des grands auditoires. Notre projet de recherche se doit dès lors de mettre en place des stratégies qui permettent d'une part de garantir un haut niveau de qualité *a priori* des épreuves et d'autre part de faire la preuve *a posteriori* de la qualité des examens standardisés.

La problématique de la « qualité », historiquement née outre-atlantique dans les milieux industriels au début du siècle, pénètre progressivement monde de la formation. Au milieu des années 90, Barnabe (1995), puisant chez une série d'auteurs spécialisés dans le domaine de la qualité, faisait le pronostic suivant : « *Tôt ou tard, notre système d'éducation devra s'intéresser sérieusement à la qualité totale. Rhodes (1992) ainsi que Glasser (1992) et Deming (1992) prétendent que l'approche qualité totale est la seule réponse aux problèmes scolaires. Comme toute autre organisation, selon Kaufman (1992), un système d'éducation utilise des ressources, développe des moyens d'action et fournit des diplômés à l'attention des clients externes. Un fait demeure, un système d'éducation entretient la promesse qu'il produira quelque chose qui sera utile à la société qui paie dans l'espoir d'obtenir des diplômés de valeur. En dernier ressort, les commentaires reçus et les pressions exercées de la part des clients externes à l'égard de ces diplômés l'obligeront à rechercher la qualité totale* ».

Du point de vue de l'utilisation des TIC en Gestion des Compétences, la littérature et les colloques se multiplient sur le sujet tant il est d'actualité dans les entreprises et mêmes les institutions publiques (réforme COPERNICUS). En septembre 2001, l'ULg a organisé un colloque sur « la GRH dans et /ou la société de l'information » qui a montré l'actualité du questionnement méthodologique et sociologique des entreprises et des chercheurs face à l'utilisation des TIC. Cependant, la réflexion pédagogique concernant l'évaluation des compétences n'y a pas été abordée. Dans la littérature française concernant la validation des acquis de l'expérience ou dans la littérature provenant des entreprises (cadres, formateurs) est largement évoquée l'instrumentation de l'évaluation des compétences : les différentes méthodes d'évaluation et leurs avantages et inconvénients y sont listés sans pratiquement jamais mentionner et encore moins analyser l'utilisation de tests standardisés. Notre hypothèse est que ce type de test (avec ses composantes qualité) est assez méconnu et que cette recherche permettrait certainement d'aider à son développement en apportant les connaissances suffisantes pour montrer l'efficacité de ce dispositif.

### 3. Rappel de l'état des connaissances des partenaires dans le domaine du projet

Dans les années '70, le Professeur Leclercq menait déjà des recherches dans le domaine des banques de questions et du testing standardisé, notamment dans le contexte de l'Ecole technique de l'armée de l'air à Saffraanberg dans le cadre de la formation des élèves pilotes et des mécaniciens de la force aérienne belge. Ces recherches ont été poursuivies dans le cadre du Service de Technologie de l'Education (STE). La reconnaissance du STE dans le domaine est internationalement reconnue (en 1992, le STE a notamment organisé le colloque international OTAN intitulé « *Item banking : Interactive Testing and Self-Assessment* »). Le Système Méthodologique d'Aide à la Réalisation de Tests (SMART) de l'ULg, issu du STE, poursuit cette voie de recherches docimologiques dans le domaine du testing standardisé. En 1996, les responsables actuels du SMART-ULg ont jetés les prémises d'un modèle de construction et de gestion de la qualité des tests standardisés (Gilles & Leclercq, 1996 ; Castaigne, Gilles & Hansen, 2001 ; Gilles, 2002), ce modèle qui sera détaillé plus loin demande actuellement à être instrumenté en procédures et outils informatisés intégrés en une plate-forme accessible à distance via le web. Cette plate-forme permettrait un transfert à grande échelle du modèle sur le terrain. Le modèle du SMART-ULg demande aussi à être adapté aux exigences des évaluations certificatives des compétences dans le monde de la formation en entreprise, cet aspect est aussi envisagé dans le cadre du partenariat HEC.

Comme décrit dans la présentation des partenaires, HEC Liège est une Haute Ecole qui, au vu de sa population grandissante, se pose quotidiennement les questions d'évaluation de qualité de grands groupes. Plusieurs enseignants ont déjà collaboré avec le SMART-ULg afin de développer des évaluations standardisées.

La recherche « *Job Employment and Orientation (JOE)* » soutenue par le FSE et coordonnée par les HEC, en collaboration avec l'ULG, a permis d'analyser l'offre du marché en terme d'outils d'aide à la gestion des compétences. De nombreux logiciels existent, proposés par des firmes privées basés sur des « *standards fermés* », et qui ont essentiellement pour objectif de créer des bases de données de compétences à partir desquelles un « manager » va pouvoir extrapoler les mobilités du personnel et la formation continue de celui-ci. Cependant, ces logiciels ne gèrent pas l'évaluation proprement dite des compétences. Ils considèrent que cela est réalisé par ailleurs.

Le logiciel développé dans le cadre de JOE ne peut prétendre à plus. Il a été expérimenté dans trois services hospitaliers, l'enseignement provincial liégeois et dans une entreprise de fret aérien. Il a été conçu sur base d'un cahier des charges réalisé avec l'aide de nombreux partenaires industriels. Cette enquête et co-élaboration de cahier des charges a démontré combien les entreprises sont peu outillées par rapport à cette gestion des compétences, et démontre encore aujourd'hui, sur le terrain, combien l'aide de « méthodologues » est indispensable pour permettre l'identification des compétences-clés d'un métier, d'un service, d'une entreprise. Par rapport à l'évaluation, le constat de lacunes technologiques et méthodologiques est encore plus important. Sous des couvertures de simplicité et de rapidité, il n'est pas rare de voir une compétence reconnue ou non à une personne sur base d'une simple croix dans une case OUI ou NON.

Dans ce contexte, une plate-forme *e-C&CQST* de création des évaluations standardisées qui, une fois élaborées, permettraient d'obtenir des résultats rapides, valides et nuancés sur les compétences des apprenants serait très appréciable.

Une telle plate-forme, accessible via Internet, intégrant toutes les étapes de construction des tests dans une seule plate-forme pourrait être utilisée en parallèle (ou couplée) avec d'autres logiciels de gestion des compétences ou avec des dispositifs d'acquisition automatique de données, par exemple une chaîne de lecture optique de marques.

#### 4. Rappel de l'innovation technologique apportée par la plate-forme *e-C&QCST*

Lorsqu'un enseignant ou un formateur est amené à créer une épreuve standardisée d'évaluation des acquis de ses apprenants il lui est actuellement difficile de faire la preuve que son test répond aux objectifs/critères de validité, de fidélité, de sensibilité, de diagnosticité, d'équité, de praticabilité et de communicabilité que nous avons rappelés au point 4.1. A l'aide des procédures docimologiques et des logiciels accessibles via le web que nous souhaitons développer dans le cadre de cette recherche, non seulement les enseignants/formateurs mais aussi les étudiants/formés et les gestionnaires des institutions/organismes de formation bénéficieront des innovations technologiques suivantes.

##### Pour les enseignants/formateurs :

- Mise à disposition d'une plate-forme de construction de tests standardisés via une démarche structurée et des procédures docimologiques permettant de maîtriser, de mesurer / contrôler et de rectifier les niveaux de qualité aux différentes étapes de réalisation des tests ;
- Gestion sécurisée et à distance des banques de questions ;
- Testing formatif ou certificatif de haute qualité avec possibilité de passation des tests à distance à l'aide de la plate-forme ;
- Possibilité de couplage avec des dispositifs d'acquisition automatique de données (par exemple une chaîne de lecture optique de marques) dans le cadre de tests « papier-crayon ».

##### Pour les étudiants/formés :

- Mise à disposition via le web d'informations certifiées à propos des procédures de testing, ces informations étant gérées en temps réel par les équipes de formateurs/évaluateurs;
- Entraînement aux épreuves à l'aide de logiciels accessibles via le web et permettant d'obtenir des informations fiables, précises et diagnostiques sur l'état des compétences des étudiants/formés ;
- Plate-forme intégrée de passation à distance des tests standardisés de certification ;
- Feedbacks de qualité (précision, exhaustivité, en temps réel, ...) après l'évaluation.

##### Pour les institutions/organismes de formation :

- Mise en place d'une plate-forme intégrée permettant de contrôler la qualité des évaluations standardisées organisées au sein de l'institution/organisme ;
- Procédures docimologiques permettant de produire des preuves de la qualité des évaluations, ces preuves pouvant servir d'arguments en termes de marketing pédagogique ;
- Pilotage et régulation de la qualité des évaluations, notamment à l'aide du module de macro-régulation ;
- Réduction des coûts liés à l'évaluation des grands groupes.

## **II. RAPPEL DE NOTRE VISION SCIENTIFIQUE DE DEPART EN CE QUI CONCERNE L'APPROCHE QUALITE EN TESTING STANDARDISE**

*Dr. Jean-Luc GILLES,  
Coordinateur, chef de projet e-C&QCST*

### **1. Pénétration des concepts « Qualité » dans les activités de testing standardisé du Système Méthodologique d'Aide à la Réalisation de Tests (SMART) de l'Université de Liège**

#### **a) Le contexte du SMART**

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 des équipes spécialisées qui proposent leur soutien aux enseignants. Beaucoup d'institutions universitaires en ont pris conscience, 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 liée aux prestations de cours et aux évaluations.

Dans et en dehors de l'institution, les TIC sont de plus en plus présentes et des spécialistes interviennent de plus en plus fréquemment lorsqu'elles sont utilisées dans les processus d'enseignement/apprentissage (réalisation des supports pédagogiques, de documents audiovisuels, de sites Internet, de tests informatisés, ...). Parallèlement, l'utilisation des TIC dans un contexte pédagogique se banalise et de plus en plus d'étudiants équipés de micro-ordinateurs, de connections au web, en exploitent les ressources.

Pratiquement toutes les grandes universités se dotent aujourd'hui de cellules spécialisées dont la mission est d'aider les enseignants. Deux exemples dans le paysage des universités francophones de Belgique : le Centre de Méthodologie des Technologies de l'Information pour l'Education (MÉTIE) à l'Université Libre de Bruxelles, l'Institut de Pédagogie Universitaire et des Multimédias (IPM) à l'Université Catholique de Louvain.

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) au besoin d'évaluer rapidement et de façon fiable de grands auditoires. Grâce au SMART, les enseignants 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 (BVE) ou à l'aide du web avec des logiciels en ligne de Testing Interactif Multimédias via Internet (TIMI).

### b) Historique du SMART

La reconnaissance officielle du SMART par les autorités académiques de l'ULg remonte à l'année académique 1998-1999, année lors de laquelle un budget fut accordé par M. le Recteur Legros pour mener à bien deux missions : (1) aider tous les enseignants ULg (quelle que soit leur faculté) dans la réalisation de leurs évaluations standardisées et (2) mettre en place le soutien logistique nécessaire à l'évaluation des enseignements de l'ULg organisée par la Commission EVALENS<sup>2</sup> mise en place par le Conseil Général des Etudes (CGE).

Historiquement, on peut dire que le SMART est né du succès rencontré par les méthodes d'évaluation des grands groupes proposées par le Centre d'Auto-Formation et d'Evaluation Interactives Multimédias (CAFEIM) basé à la Faculté de Psychologie et des Sciences de l'Education (FAPSE) de l'ULg.

Le CAFEIM<sup>3</sup> fut créé en 1990 par le Professeur Leclercq du Service de Technologie de l'Education (STE) avec le soutien de la FAPSE et en collaboration avec les professeurs Crahay (Service de Pédagogie Expérimentale), Henry (Service de Développement et Evaluation de Programmes Pédagogiques) et Reginster-Haneuse (Ecole de Santé Publique). Lors de l'année académique 1994-1995, le CAFEIM devint une structure officielle de la FAPSE-ULg. Sous l'impulsion du Doyen de la FAPSE, à l'époque le Professeur De Keyser, le centre fut chargé d'aider les enseignants à réaliser leurs examens en mettant à leur disposition un dispositif ayant recours aux techniques d'évaluation informatisées, notamment par Questions à Choix Multiples avec Solutions Générales Implicites<sup>4</sup> (QCM-SGI) et degrés de certitude (D. Leclercq, 1975, 1983, 1986, 1993). Ces interventions avaient lieu en priorité au premier cycle, là où les étudiants sont les plus nombreux : plus de 400 en première année (parfois plus de 600 lorsque des sections sont regroupées) et plus de 200 en deuxième année.

La mise en œuvre de ce service d'aide à la réalisation d'évaluations standardisées fut rendue techniquement possible grâce à des crédits de recherche octroyés par le FNRS<sup>5</sup> au STE en vue d'équiper ce service d'un dispositif complet de Lecture Optique de Marques. Cette chaîne LOM fut mise à la disposition du CAFEIM qui fut ainsi en mesure de répondre efficacement aux nombreuses demandes. A l'époque, l'augmentation des effectifs dans le premier cycle d'études amena des changements dans les pratiques d'évaluation chez des enseignants confrontés à des groupes d'étudiants de plus en plus nombreux.

<sup>2</sup> Président de la Commission EVALENS : Prof. Claude Houssier.

<sup>3</sup> Le CAFEIM (Président : Prof. Dieudonné Leclercq, directeur de 1995 à 2000 : Jean-Luc Gilles) s'est fixé pour objectifs :

- de promouvoir l'apprentissage autonome et coopératif recourant aux multimédias ;
- de faciliter les évaluations de grands groupes d'étudiants en recourant aux technologies appropriées (QCM, lecture optique, etc.) ;
- de permettre les animations et/ou les évaluations interactives individuelles ou collectives (théâtre électronique) ;
- de développer une activité de recherche scientifique en matière d'application des technologies de l'information à l'apprentissage, l'enseignement et l'évaluation ;
- de favoriser des activités de conception et de réalisation de produits de formation.

<sup>4</sup> Les Solutions Générales Implicites (Leclercq, 1986) 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 correcte), Absurdité (il y a une contrevérité dans l'énoncé à *dénoncer en priorité* !).

<sup>5</sup> Fonds National de la Recherche Scientifique

Vu la taille des auditoires, le recours aux épreuves standardisées avec questions à choix multiple devint incontournable pour certains. Sur 61 épreuves organisées en première et deuxième candidatures en 1997-1998 à la FAPSE, 24 (39,3%) étaient de type QCM (Debry & al., 1999) et corrigées à l'aide du dispositif LOM, à l'exception d'une seule épreuve<sup>6</sup> qui avait recours à un système de testing informatisé *WINCHECK*<sup>7</sup> où les questions étaient posées et les réponses traitées via les ordinateurs de la salle multimédias du CAFEIM (les étudiants y disposent chacun d'un ordinateur et sont évalués par cohortes de 24). Dès le démarrage de ces activités, des professeurs donnant des cours dans le second cycle ou provenant d'autres facultés de l'ULg, marquèrent aussi leur intérêt pour cette aide méthodologique et logistique dans la réalisation des examens standardisés. En deux années académiques, 1994-1995 et 1995-1996, c'est plus d'une septantaine d'évaluations standardisées qui furent ainsi organisées avec l'aide du CAFEIM (Gilles, 1996a; 1996b).

L'accroissement des demandes et le caractère de plus en plus « interfacultaire » de celles-ci (elles émanent des 8 facultés de l'ULg) provoqua une prise de conscience en ce qui concerne la nécessité de subventionner cette réponse aux besoins des professeurs de toute l'ULg par un financement provenant des autorités académiques. Il faut dire que jusqu'en 1998 l'aide fournie était financée par des crédits de la FAPSE (qui lui octroyait aussi un assistant facultaire) et rendue possible par des apports en équipement et en personnel technique provenant du STE. Devant le succès rencontré et l'accroissement des demandes, cette situation mettait en péril l'équilibre budgétaire du CAFEIM (structure interne à la FAPSE) et le fonctionnement interne du STE. Le problème fut résolu par la création d'une nouvelle cellule à vocation interfacultaire, le Système Méthodologique d'Aide à la Réalisation de Tests (SMART), qui fut financée en 1998 par les autorités académiques grâce à un budget octroyé par M. le Recteur Legros.

On le voit, la naissance du SMART<sup>8</sup> fut possible grâce aux efforts conjugués du Service de Technologie de l'Education (STE), de la Faculté de Psychologie et des Sciences de l'Education (FAPSE), du Centre d'Auto-Formation et d'Evaluation Interactives Multimédias (CAFEIM) ainsi que des autorités académiques de l'ULg (Rectorat). Mais avant tout, ce qui explique l'émergence et le succès de cette cellule interfacultaire, c'est le fait qu'elle réponde à des besoins rencontrés par de plus en plus de professeurs confrontés aux mêmes problèmes méthodologiques et logistiques que posent les évaluations des grands groupes d'étudiants. Les services proposés par le SMART ne se sont jamais cantonnés à l'unique phase de correction des examens. En fait, dès le départ, l'aide aux enseignants consista en une série de services docimologiques qui forment un tout cohérent et s'inscrivent dans un cycle de gestion des épreuves standardisées que nous détaillerons plus loin. Voici en bref les différents types de services proposés. Il s'agit d'aider les professeurs à :

- choisir le(s) type(s) de questionnements, la(les) méthode(s) de testing la(les) plus appropriée(s) ;
- gérer la(les) banque(s) de questions ;
- analyser la qualité *a priori* des épreuves (tests formatifs ou examens) ;
- entraîner les étudiants aux procédures d'évaluation ;
- préparer la logistique des épreuves ;
- corriger les tests à l'aide de procédures informatisées ;

<sup>6</sup> L'examen du cours de Psychologie développementale, 1<sup>ère</sup> candidature, Prof. Michel Born.

<sup>7</sup> *WINCHECK* est un logiciel d'évaluation interactive multimédia développé par le STE.

<sup>8</sup> Responsable académique : Prof. Dieudonné Leclercq, directeur : dr. Jean-Luc Gilles.

- analyser la qualité *a posteriori* des questions ;
- mettre en place des solutions pour rectifier les épreuves lorsque des problèmes sont détectés ;
- fournir des feedbacks individualisés aux étudiants relatifs à leurs performances et aux performances de l'épreuve ;
- recueillir les avis des étudiants sur la qualité des épreuves.

En plus du dispositif de la lecture optique de marques, le SMART a été doté d'autres dispositifs technologiques que nous présenterons plus loin : les Boîtiers de Vote Electroniques (BVE) et le système de Testing Informatisé Multimédia via Internet où les étudiants répondent aux questions via un ordinateur (TIMI). Plus récemment le SMART a conçu un logiciel en ligne en vue de fournir une aide à distance aux enseignants dans le cadre de la correction de leurs tests. Ce service à distance est intitulé Traitements Via Internet (TVI).

### c) Facteurs de mutation de l'université depuis les années 60

Le SMART se soucie de la qualité des épreuves qui sont proposées via son aide méthodologique. Cette préoccupation relative à la qualité des services offerts aux étudiants et aux enseignants s'insère dans un mouvement qui depuis le début des années '90, touche les institutions universitaires belges engagées dans une démarche d'évaluation de la qualité (Kaufman & al., 1995 ; Segers & Dochy, 1996). Ce mouvement participe lui-même à une tendance mondiale d'introduction du management de la qualité dans la sphère des activités universitaires (Ramsden, 1991 ; Nightingale & O'Neil, 1994 ; Zink & Schmidt, 1995).

Les pratiques évaluatives de l'université n'échappent évidemment pas à cette tendance récente d'évaluation de la qualité, c'est ce que soulignent Blais & al. (1997) : « *Dans le contexte social actuel les établissements d'enseignement supérieur sont de plus en plus souvent amenés à faire la démonstration de la qualité de la formation dispensée. En ce sens, l'évaluation des apprentissages, comme mécanisme de régulation et de contrôle exige une attention particulière. Par ailleurs, la régulation et le contrôle des apprentissages doivent prendre en considération un souci de plus en plus net d'assurer l'équité et la transparence des pratiques d'évaluation* ». Conscient de cette ligne de force actuelle, le SMART a entamé une réflexion sur la mise en place de méthodes de gestion de la qualité des épreuves standardisées en vue de satisfaire au mieux les enseignants et les étudiants. Notre réponse se situe dans un contexte universitaire qui s'est fortement modifié en une génération.

En effet, l'institution universitaire a considérablement évolué depuis les années 60. Divers facteurs expliquent les mutations de l'université. Dupont & Ossandon (1994, p.13) citent notamment la diversification de l'origine sociale des étudiants : « *... alors que les universitaires d'antan provenaient de filières secondaires sélectionnées, on observe que l'origine scolaire secondaire appartient à un éventail extrêmement varié de formations* ». Ces auteurs font l'hypothèse d'une diminution de la motivation en lien avec cette diversification sociale : « *la motivation sans doute plus affirmée de l'élite d'antan s'accompagne maintenant de groupes d'étudiants plus nombreux qui entreprennent des études universitaires sans véritable projet de carrière* ». A cette diversification sociale on peut aussi ajouter une tendance nouvelle de diversification des âges de ceux qui viennent se former à l'université.

L'évolution rapide des sciences et des techniques entraîne la nécessité d'une formation continue des universitaires : aujourd'hui l'offre des formations en troisième cycle est en pleine explosion (surtout dans les domaines en pleine mutation). On remarque également la création de passerelles de plus en plus nombreuses entre les formations, favorisant ainsi encore plus de diversité dans les publics. Et à ce constat de la diversification des publics, s'ajoute celui de la massification. Récemment à l'Université de Liège, des amphithéâtres pouvant contenir plus de 600 étudiants ont été construits. Dès lors, la pédagogie des grands groupes s'impose par elle-même et comme le soulignent Dupont & Ossandon (1994, p.15) : *« par la force des choses, la pédagogie universitaire, prise sous l'angle des méthodologies utilisées, est de plus en plus inadéquate ; la plupart des universités européennes sont confrontées à une pédagogie des grands groupes à laquelle la majorité des enseignants n'ont pas été sensibilisés ; on est loin d'un taux d'encadrement de 1 enseignant pour 15 à 20 élèves rencontré dans les prestigieuses universités américaines »*.

La pédagogie des grands groupes est souvent connotée négativement, il en va de même pour les évaluations standardisées des grands auditoires. Une cellule comme le SMART se doit dès lors de mettre en place des stratégies qui permettent d'une part de garantir un niveau de qualité *a priori* des épreuves et d'autre part de faire la preuve *a posteriori* de la qualité des examens standardisés. Elle doit également veiller à faire connaître les niveaux de qualité atteints et mettre en œuvre des procédures docimologiques de rectification des épreuves lorsque des problèmes se posent (nous verrons plus loin quelles procédures ont été mise en place dans le cadre du SMART). La mise en œuvre de canaux de communication entre les enseignants, les étudiants et le SMART est nécessaire pour maintenir la confiance dans les procédures proposées. La preuve de la qualité et la communication de cette preuve (qui permet sa discussion) constituent à notre avis deux socles indispensables à la gestion de la qualité des épreuves standardisées universitaires.

Revenons aux facteurs de mutation de l'université, Romainville & Boxus (1998, p.23) invoquent d'autres facteurs se situant au niveau de la conception même de l'université. Ils font le constat du passage d'une université instrument de justice sociale dans les années 60-70 *« ... fièrement indépendante des milieux économiques ... »* à une université soumise aux exigences du monde industriel dans les années 80, période où l'état des finances publiques impose une sacro-sainte rigueur budgétaire tout en confiant à l'université une nouvelle mission, celle de favoriser la reconversion économique : *« ... on exige que l'université participe activement au développement de la compétitivité du pays »*. D'après les auteurs, l'introduction de préoccupations nouvelles comme l'évaluation de la qualité au sein de l'université serait à mettre en lien avec cette volte-face de l'institution se rapprochant des milieux économiques. Un autre facteur évoqué est lié à l'idée que la mondialisation n'est pas qu'économique : *« ... il existe des solutions internationales aux problèmes que rencontrent les universités d'un pays donné. Ainsi, les Pays-Bas ont largement imité les procédures en cours aux Etats-Unis. »*. On peut aussi ajouter qu'en plus de la concurrence intra-nationale, une concurrence internationale s'installe entre les universités. Les universités *« captent »* des publics provenant d'horizons de plus en plus éloignés dépassant les frontières nationales. C'est aussi une nouvelle tendance que soulignent Dupont & Ossandon (1994, p.14), *« ... de plus en plus, l'enseignement passera du monoculturel et du monolingue au pluriculturel et au plurilingue en raison d'une part des programmes Erasmus et, d'autre part, de la libre circulation des personnes »*. Le mélange dans les amphithéâtres d'étudiants *« du cru »* avec des étudiants provenant d'autres institutions universitaires posera tôt ou tard le problème de l'uniformisation et de la qualité des programmes de cours dispensés dans les différentes universités.

Ces facteurs de mutation liés à la diversification et la massification des publics, au rapprochement avec les milieux industriels et à l'internationalisation de l'offre de formation ont amené de façon progressive la pénétration au sein de l'université de concepts nouveaux liés à la problématique de « l'évaluation de la qualité » historiquement née outre-atlantique dans les milieux industriels au début du siècle. Au milieu des années 90, Barnabe (1995), puisant chez une série d'auteurs spécialisés dans le domaine de la qualité, faisait le pronostic suivant : *« Tôt ou tard, notre système d'éducation devra s'intéresser sérieusement à la qualité totale. Rhodes (1992) ainsi que Glasser (1992) et Deming (1992) prétendent que l'approche qualité totale est la seule réponse aux problèmes scolaires. Comme toute autre organisation, selon Kaufman (1992), un système d'éducation utilise des ressources, développe des moyens d'action et fournit des diplômés à l'attention des clients externes. Un fait demeure, un système d'éducation entretient la promesse qu'il produira quelque chose qui sera utile à la société qui paie dans l'espoir d'obtenir des diplômés de valeur. En dernier ressort, les commentaires reçus et les pressions exercées de la part des clients externes à l'égard de ces diplômés l'obligeront à rechercher la qualité totale ».*

En parallèle avec l'apparition des préoccupations relatives à l'évaluation de la qualité, on assiste aussi aujourd'hui à l'émergence d'un véritable « marketing universitaire ». Les institutions soignent leur image de marque et cherchent à se positionner sur le marché de l'offre des formations en mettant en avant leurs atouts respectifs. Dans ce contexte, la mise en évidence de la qualité des enseignements et services offerts aux étudiants constitue un argument susceptible d'alimenter les politiques de communications vers ceux qui sont les futurs « clients » de l'université. La qualité des examens auxquels les étudiants sont soumis pourrait bien dans un avenir proche constituer un nouvel argument des institutions d'enseignement supérieur vis à vis de leurs publics cibles.

On le voit, le contexte des années 90 était mûr pour qu'une cellule spécialisée dans les problèmes d'évaluation des grands groupes voie le jour. Tout comme le recours à des méthodes et des dispositifs de testing utilisant les Technologies de l'Information et de la Communication (TIC) est incontournable pour répondre efficacement aux problèmes d'évaluation standardisée des grands auditoires ; la mise en œuvre d'une gestion qualité des épreuves est essentielle pour garantir un service optimal adapté aux exigences des enseignants et étudiants. Dans ce contexte, c'est aussi assez « naturellement » qu'une réflexion sur la problématique de la qualité des épreuves standardisées universitaires s'est construite au sein du SMART.

#### **d) Définir le concept de « Qualité »**

Produire une définition de la qualité qui soit acceptable par une large majorité semble constituer un exercice difficile voire impossible. Pour le Conseil de l'Education et de la Formation (CEF)<sup>9</sup> : *« ...la qualité est un concept complexe, qui diffère selon le produit, selon les attentes des utilisateurs, selon les processus et selon les résultats ».* Si on se réfère à la définition officielle donnée par la norme ISO 8402<sup>10</sup>, la qualité : *« est l'ensemble des propriétés et caractéristiques d'un produit ou service qui lui confèrent l'aptitude à satisfaire des besoins exprimés et implicites ».*

<sup>9</sup> Conseil de l'Education et de la Formation, CEF/avis, n°52/9, janvier 1998.

<sup>10</sup> Institut Belge de Normalisation (IBN), 1990.

Il est heureux que le document ISO officielle complète cette définition d'une série de notes pour en préciser le sens. Ces notes mettent l'accent sur des nuances importantes :

- la nécessité d'identifier et de définir les besoins implicites ;
- la traduction des besoins en propriétés et caractéristiques (ce qui permet d'envisager des comparaisons besoins-produits). Ces besoins pouvant, entre autres, comporter des aspects de sûreté, de fiabilité, de « maintenabilité » ;
- l'adaptation des spécifications en fonction de l'évolution des besoins ;
- la nécessité d'accompagner le terme « qualité » d'un qualificatif lorsqu'il s'agit de l'utiliser pour exprimer un degré d'excellence dans un sens comparatif : « *qualité relative lorsque les produits ou services sont classés par degré d'excellence* » ou « *niveau de qualité et mesure de qualité lorsque des évaluations techniques précises sont effectuées quantitativement* » ;
- l'influence des phases du processus de réalisation : « *la qualité d'un produit ou service est influencée par de nombreuses phases d'activités interdépendantes, telles que la conception, la production ou le service après-vente et la maintenance.* » ;
- l'influence des apports à la qualité des différentes phases du processus réalisation est aussi mis en rapport avec le concept de « *boucle de la qualité* » ou « *spirale de la qualité* » lui-même défini un peu plus loin dans la norme ISO 8402 (3.3 p.6) comme étant le « *Modèle conceptuel des activités interdépendantes qui exercent leur influence sur la qualité d'un produit ou service tout au long des phases qui vont de l'identification des besoins jusqu'à l'évaluation de leur satisfaction* ». Dès lors, il est possible d'identifier l'apport en qualité dû à une phase donnée, par exemple la « *qualité due à la conception* » ou la « *qualité due à la mise en œuvre* ».

Ainsi nuancée, la définition ISO 8402 de la qualité est beaucoup moins éloignée qu'il n'y paraît de prime abord de la définition de la Fédération des Etudiants Francophones (FEF)<sup>11</sup> : « *La qualité n'existe pas une fois pour toutes. Elle relève de la gestion de projet. Un projet se définit avec l'ensemble des acteurs concernés, est ensuite mis en application... avant d'être évalué. Après évaluation, le projet est reprécisé (redéfini) et ainsi de suite... La qualité s'inscrit dans un processus progressif infini qui fait que sa limite (l'excellence) n'est jamais atteinte.* »

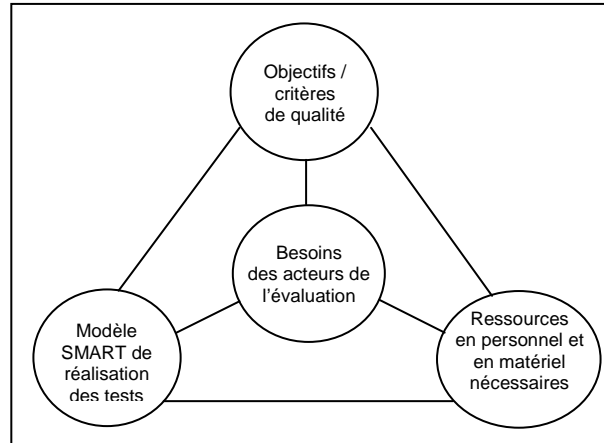
La définition ISO 8402 et celle de la FEF se rejoignent dans l'idée que la qualité peut être constamment améliorée, qu'elle doit être envisagée comme faisant partie d'un processus cyclique où après chaque cycle de réalisation mais au sein de chaque phase de sa construction, le produit, le service, l'examen (dans le cadre des épreuves standardisées universitaires) s'améliorent pour tendre vers une qualité maximale (l'excellence) qui n'est jamais atteinte. C'est le concept de « *spirale de la qualité* ». Ce concept est intégré dans notre modèle de gestion des évaluations universitaires standardisées que nous présenterons plus loin. Mais avant, clarifions quels sont les objectifs du SMART en ce qui concerne la « *qualité des épreuves standardisées* ».

---

<sup>11</sup> Fédération des Etudiants Francophones, *Les modules capitalisables*, Bruxelles, 17 novembre 1997.

## 2. Critères de qualité des évaluations standardisées

*Les recommandations de la norme internationale ISO 9004-2<sup>12</sup> (Gestion de la qualité et éléments de système qualité - Lignes directrices pour les services) mettent en évidence les facteurs clés d'un système qualité. Dans le contexte du SMART ces facteurs sont illustrés par le schéma ci-contre. Au centre : les « Besoins des acteurs de l'évaluation » (évaluateurs, évalués et de façon plus large tous les intervenants dans une situation de testing standardisé), destinataires des services proposés par le SMART et points de convergence des autres facteurs clés. A la base : d'une part le modèle de construction en spirale de qualité des examens que nous utilisons dans le cadre du « Modèle SMART de réalisation des examens » et, d'autre part, les « Ressources en personnel et en matériel nécessaires au SMART » pour la mise en œuvre du processus de réalisation des épreuves. En haut du schéma nous avons placé les « Objectifs / critères de qualité », c'est d'eux dont il sera question dans cette partie.*



Succinctement, en ce qui concerne le SMART, notre politique de qualité est orientée vers la mise en place d'évaluations qui devraient offrir aux enseignants et étudiants des garanties de :

- **Validité**, les scores des apprenants doivent représenter ce que l'enseignant / le formateur veut mesurer, permettre des inférences solides (validité de *construct* ou théorique), et couvrir les aspects importants des apprentissages (validité de contenu) ;
- **Fidélité**, la note (d'une épreuve corrigée) classée dans une catégorie donnée doit l'être de la même façon si elle est traitée par d'autres correcteurs (concordance interjuges) ou/et à un autre moment (consistance intrajuge) ;
- **Sensibilité**, la mesure doit être précise, refléter des phénomènes subtils ;
- **Diagnosticité**, les résultats doivent permettre le diagnostic précis des difficultés d'apprentissage (et idéalement de leurs causes), des processus maîtrisés, de ceux qui ne le sont pas ;
- **Praticabilité**, la réalisation des évaluations doit être faisable endéans des délais raisonnables et à l'aide des ressources en personnel et en matériel disponibles et proportionnées à l'importance relative des enjeux ;
- **Équité**, tous les étudiants doivent être traités de façon juste, en principe de la même manière (standardisation) ;
- **Communicabilité**, les informations non confidentielles relatives au déroulement du processus doivent être communiquées et comprises par les partenaires (enseignants/formateurs, évalués, équipe SMART, ...) engagés dans la réalisation des épreuves.
- **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.

<sup>12</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.

Envisageons maintenant de manière plus détaillée ces différents objectifs « qualité » à atteindre dans la réalisation des examens standardisés.

### a) La validité

Selon De Landsheere (1980, p. 178), valider une épreuve consiste à « *apporter la preuve (...) que l'examen fournit une évaluation correcte de ce qu'il prétend mesurer ou prédire* ». En d'autres termes, il s'agit de se poser la question : les questions posées aux étudiants correspondent-elles bien à ce que l'examineur veut mesurer ? Répondre à cette question implique que la clarté ait été préalablement faite sur ce qui doit être mesuré, que les questions de l'examen soient clairement mises en lien avec les objectifs du cours. On parlera alors de validité de contenu. Nous allons voir qu'il existe différentes formes de validité et qu'on s'accorde aujourd'hui à en définir trois grands types.

#### *Evolution des conceptions en matière de validité*

Le concept de validité a considérablement évolué depuis un demi siècle. Laveault & Gregoire (1997, p. 189) signalent (nous soulignons) : « *Au début des années 50 (Messick, 1988, pp. 18-19), la validité était envisagée de manière relativement morcelée. Ainsi, les 'Technical Recommendations' de l'American Psychology Association (1954) se limitaient à codifier des types de validité (de contenu, prédictive, concomitante et conceptuelle). (...) Il faut attendre les années 70 pour qu'un effort important soit réalisé dans le sens d'une intégration des différents types de validité. L'aboutissement de cet effort est manifeste dans les 'Standards for Educational and Psychological Testing' publiés conjointement par l'American Psychological Association et l'American Educational Research Association en 1985. Les 'Standards' constituent aujourd'hui une référence incontournable pour les spécialistes de la mesure en psychologie et en éducation. (...) la validité est présentée comme un concept unitaire se rapportant non au test lui même mais aux inférences faites à partir des résultats de celui-ci. Dans cette perspective... Seules sont valides les inférences en faveur desquelles suffisamment d'arguments et de données empiriques ont pu être rassemblés. (...) Suivant cette conception de la validité, la validation est un processus toujours continu d'accumulation de preuves. Les types de validité, définis dans les ouvrages des années 50 et 60 sont aujourd'hui envisagés comme moyens de validation servant à rassembler des arguments en faveur de telle ou telle inférence.* »

Ainsi, les 'standards' de 1985 distinguent trois grands types de validation : (1) la validation relative au contenu, (2) la validation en référence à un critère externe et (3) la validation en référence à un concept ou un modèle théorique.

#### *La validation relative au contenu*

Il s'agit dans ce cas de demander l'avis des experts du contenu, la question qui leur est posée étant « Les items de l'épreuve sont-ils représentatifs du concept ou du domaine évalué ? ». Une épreuve sera donc jugée valide du point de vue du contenu si les arguments (subjectifs) des experts s'accordent sur le fait que le test de connaissances qui ambitionne de faire le bilan des acquis liés à un enseignement couvre réellement les aspects importants de cet enseignement. De Landsheere (1979, p. 291) introduit l'idée d'une sélection plus rationnelle des aspects d'un enseignement qui devront être évalués dans une épreuve (nous soulignons) : « *L'appréciation de l'importance repose soit sur un jugement de valeur, soit sur un raisonnement (...) dans quelle mesure tel apprentissage est-il nécessaire pour accéder à tel autre, jugé important ? (C'est pour-*

quoi on parle parfois dans ce dernier cas de validité rationnelle ou logique) ».

De Landsheere signale également qu'on peut considérer qu'un test manque de validité de contenu s'il n'explore pas un éventail suffisant de comportements, par exemple s'il ne fait appel qu'à la mémoire (*op. cit.* p. 291). C'est tout le problème de la rupture entre enseignement et examen qui est posé. De Landsheere (1980, p.29) évoque ainsi le cas du professeur de chimie dont 4 élèves sur 22 obtenaient la moitié des points à l'examen : « *Le professeur ne faisait que deux ou trois interrogations par an. Sur les trois questions d'examen, deux impliquaient interprétation et transfert, démarches qui, au niveau de difficulté où le maître se plaçait, nécessitaient une compréhension profonde des phénomènes et un entraînement, long et finement contrôlé, à la solution de problèmes originaux. Une discussion amicale avec le jeune professeur en question révéla qu'il se croyait obligé de traiter en détail de tous les points du programme, qu'il n'avait pas conscience de la nécessité de l'entraînement systématique et au moins semi-individualisé à la solution de problèmes originaux et que, de toute façon, il ne possédait pas les notions psychopédagogiques de base nécessaire à la conduite d'un tel entraînement* ».

Nous décrivons plus loin, dans le cadre du cycle en « spirale de qualité » de réalisation des épreuves universitaires sommatives standardisées, une procédure de validation rationnelle relative au contenu à couvrir par un examen. Cette procédure consiste à identifier dans un premier temps, d'une part les Points Enseignés (PE) et d'autre part les Catégories de Performances (CP). On crée ainsi une liste de binômes unissant à chaque fois un point enseigné avec une catégorie de performances : [PE x CP] et on en détermine quel sont les binômes essentiels. Ensuite lors de la mise en forme de l'épreuve les Modalités de Questionnement (MQ) sont définies. Ces MQ sont ensuite associées aux Catégories de Performances (CP) mises en évidence à l'étape précédente. On obtient alors une autre série de binômes [CP x MQ]. Enfin, pour chaque Point Enseigné mis en évidence précédemment, on peut alors concevoir un trinôme [PE x CP x MQ]. Chaque trinôme est donc formé d'un Point Enseigné, d'une Catégorie de Performance et d'une Modalité de Questionnement. Cela donne par exemple le trinôme suivant : « *Identification sur l'image d'une échographie [CP] d'un kyste ovarien [PE] évaluée à l'aide d'une QCM-SGI à livre ouvert [MQ]* ».

Ceci dit, et malgré le caractère structuré de cette démarche, il faut bien reconnaître que déterminer quels sont les binômes essentiels qui devront être évalués relève finalement du jugement de valeur, de croyances en l'importance des PE, et, en dernier ressort, c'est le bon sens qui fini par emporter la décision. Laveault & Gregoire (*op. cit.*) font remarquer à propos de la validation relative au contenu : « *Par définition cette modalité de validation des tests est subjective. Toutefois, si elle respecte une méthodologie rigoureuse, elle permet d'arriver à des conclusions solides qui pourront trouver confirmation dans des recherches empiriques ultérieures* ».

Si il est fait appel à des juges non entraînés, pas forcément experts, pour évaluer si les questions d'une épreuve couvrent bien un domaine donné et si ces juges n'utilisent pas de méthodologie rigoureuse on parlera alors plutôt de validation apparente (*face validity*). Cette procédure est donc moins scientifique mais peut dans certains cas être utile pour orienter la façon dont les questions devraient être posées et faciliter ainsi l'acceptation d'épreuves par les utilisateurs. De Landsheere (1979, p. 290) relate l'exemple d'un test soumis à un groupe de candidats aviateurs : « *... Ainsi, R. Thorndike et E. Hagen observent qu'un groupe de candidats aviateurs sera plus disposé à accepter un test d'arithmétique dont les problèmes portent sur la vitesse du vent ou sur la consommation de carburant qu'un test faisant porter les mêmes types de problèmes sur l'agriculture* ».

### La validation en référence à un critère externe

Laveault & Gregoire (1997, p. 190) distinguent deux types de validation en référence à un critère externe, d'une part la validation concomitante et d'autre part la validation prédictive.

Dans le cas de la validation concomitante, il s'agit de vérifier si les corrélations entre les scores à l'examen et les mesures prises comme critère de référence sont suffisantes. On examinera par exemple la corrélation entre les scores à un test et ceux obtenus à une autre épreuve ou une série d'autres épreuves considéré(s) comme une référence solide.

Dans le cas de la validation prédictive l'idée est de vérifier si le pronostic formulé par le test se vérifie ou non. Selon Laveault & Gregoire (*op. cit.*) « *La validation prédictive consiste, quant à elle, à évaluer la qualité des prédictions faites sur base des scores au test, dans ce cas, le critère est la mesure de ce qui a été prédit. Par exemple, la validation d'un test d'admission consistera en la comparaison des scores au test et des résultats obtenus à la fin d'un programme d'étude* ».

### La validation en référence à un concept ou un modèle théorique

Les scores des étudiants doivent permettre des inférences solides (validité de *construct* ou théorique ou conceptuelle). Les épreuves reposent sur des modèles implicites ou explicites de la réalité qu'elles sont sensées évaluer. Ce sont ces modèles qui permettent de donner du sens aux résultats des tests.

Voici un exemple de modèle utile à l'interprétation des résultats. Il s'agit du modèle qui a présidé à la conception de « double check » (Leclercq, 1993), une procédure d'évaluation interactive, qui consiste à poser une question en deux volets *prim* et *bis*. L'étudiant reçoit une première (*prim*) question (QCM-SGI) où la réponse correcte attendue peut, par exemple, être « 8. Manque de données dans l'énoncé ». Après avoir répondu, l'étudiant reçoit la réponse puis la deuxième partie de la question (*bis*), par exemple : « quelle donnée manque ? ». Suit à nouveau une série de propositions. Les performances des étudiants se présentent alors selon différents cas de figure qui peuvent ensuite donner lieu à des procédures de remédiation adaptées selon le diagnostic fourni par le modèle.

	<b>Volet <i>prim</i> (analyse)</b>	<b>Volet <i>bis</i> (compréhension)</b>	<b>Diagnostic :</b>
<b>Compétence totale</b>	Réussite	Réussite	Analyse et compréhension correctes du problème.
<b>Compétence partielle de type a</b>	Echec	Réussite	Manque de vigilance (mais compréhension).
<b>Compétence partielle de type b</b>	Réussite	Echec	Incompréhension du problème. La solution choisie dans le volet bis peut indiquer l'erreur de raisonnement.
<b>Incompétence</b>	Echec	Echec	Les solutions choisies dans la partie <i>prim</i> et <i>bis</i> peuvent donner des indications quant aux causes de l'échec.

### *Trois techniques qui renforcent la validité de contenu des épreuves standardisées universitaires*

Précédemment nous avons évoqué la nécessité de distinguer les QCM classiques d'un autre type de QCM dont les modalités améliorent considérablement la qualité de cette forme de questionnement fermé.

La plupart du temps, les enseignants et les formateurs 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 qui permettent de clarifier ces processus mentaux. Ainsi la taxonomie de Bloom (1969) souvent utilisée pour classer différents niveaux de performances, est utile pour distinguer : (1) la connaissance ou 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 détection de problèmes et les classifications, (5) la synthèse, c'est-à-dire expression et (re)formulations, (6) l'évaluation ou le jugement sur base de critères personnels.

Les Questions à Choix Multiple (QCM) habituelles permettent de mesurer les niveaux (1) connaissance, (2) compréhension et (3) application. Mais, dans le cadre d'épreuves standardisées destinées à de grands groupes, lorsqu'il s'agit d'évaluer (4) l'analyse c'est à un autre type QCM, les QCM-SGI (que nous allons aborder ci-dessous) qu'il faut faire appel.

Nous présenterons aussi plus loin la technique des degrés de certitude (Leclercq, 1983, 1993) qui constitue une réponse proposée par le SMART en ce qui concerne l'évaluation du niveau taxonomique « (6) l'évaluation ou le jugement sur base de critères personnels » à propos de la matière enseignée.

Enfin, nous terminerons en abordant le questionnement à livre ouvert. Cette technique est en effet plus adéquate lorsque l'examineur souhaite mesurer le niveau (2) de compréhension et non le niveau (1) de connaissance.

#### *Les QCM-SGI*

Leclercq (1993, p. 210) a montré qu'il était possible d'évaluer (4) l'analyse de façon systématique en ayant recours aux Questions à Choix Multiple avec Solutions Générales Implicites (QCM-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 contrevérité dans l'énoncé à dénoncer en priorité !). L'auteur souligne (*op. cit.*, p. 211) que la performance demandée à l'étudiant dans le cadre des QCM-SGI ne relève plus de la simple (re)connaissance, mais de processus mentaux plus élevés dans la taxonomie de Bloom : « *These type of question force the student to consider the relevance of the data or of the question themselves, i.e. to understand the problem and analyse the way it is stated instead of limiting him/herself at plain knowledge and application levels (in the terms of Bloom's taxonomy of cognitive objectives). We have adopted this principle, making it systematic, i.e. applying it in all our questions* ».

Gilles et Melon (2000) ont procédé à une comparaison des résultats au test de Compréhension d'un Texte Scientifique<sup>13</sup> (CTS) de mars 1998 avec un test de Maîtrise du Français (MF) comportant 80 questions réparties en 4 sous-tests : orthographe (ORTH), vocabulaire (VOCA), syntaxe (SYNT) et compréhension (COMP) de 20 questions chacun.

Corrélations significatives marquées en gras à  $p < ,05$  (N=199)

	CTS		MF				Partiels 1 à 5 tests		
	17 Q. SGI	9 Q. HABI	80 Q.	20 Q. ORTH	20 Q. VOCA	20 Q. SYNT		20 Q. COMP	
<b>CTS</b>	26 Q.	<b>,91</b>	<b>,75</b>	<b>,60</b>	<b>,35</b>	<b>,52</b>	<b>,59</b>	<b>,48</b>	<b>,55</b>
	17 Q. SGI		<b>,41</b>	<b>,54</b>	<b>,29</b>	<b>,50</b>	<b>,50</b>	<b>,44</b>	<b>,58</b>
	9 Q. HABI			<b>,47</b>	<b>,31</b>	<b>,34</b>	<b>,50</b>	<b>,36</b>	<b>,29</b>
<b>MF</b>	80 Q.				<b>,75</b>	<b>,83</b>	<b>,88</b>	<b>,73</b>	<b>,39</b>
	20 Q. ORTH					<b>,42</b>	<b>,54</b>	<b>,39</b>	<b>,23</b>
	20 Q. VOCA						<b>,69</b>	<b>,48</b>	<b>,38</b>
	20 Q. SYNT							<b>,57</b>	<b>,33</b>
	20 Q. COMP								<b>,30</b>

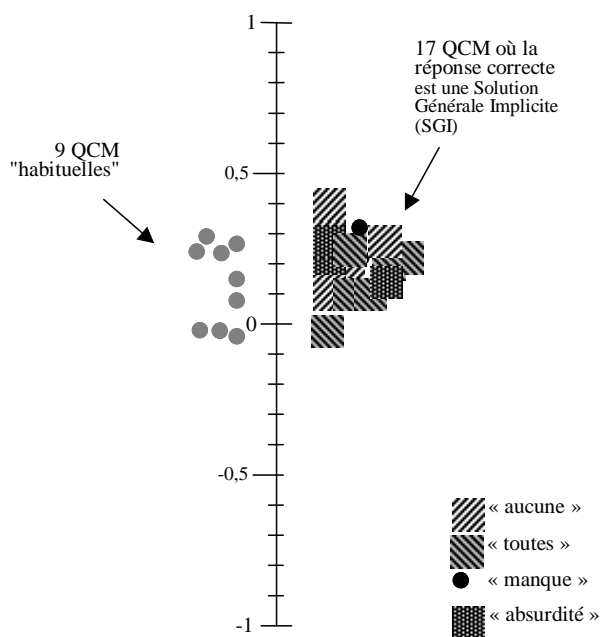
Tableau 2 : corrélations des scores aux partiels et des scores avec degrés de certitude aux tests MF & CTS

Le test CTS comportait quand à lui 27 questions réparties en 17 QCM dont les réponses correctes étaient des Solutions Générales Implicites (SGI) et 9 autres QCM dont les réponses correctes étaient habituelles (HABI).

Les résultats à ces tests ont aussi été comparés avec les performances académiques des étudiants aux examens partiels de janvier (dernière colonne du tableau).

Les corrélations présentées dans le tableau ci-dessus montrent que ce sont les performances au test CTS qui sont les plus corrélées ( $r = .55$ ) avec les résultats obtenus aux examens partiels de janvier même lorsqu'on le compare au test de Maîtrise du Français (MF) qui comporte cependant plus de questions.

Au sein du test CTS, les performances aux 17 questions dont la réponse attendue était une SGI sont encore plus corrélées avec les partiels ( $r = .58$ ) mais ces questions étaient aussi plus nombreuses dans le test. Dès lors, qu'en est-il lorsqu'on compare question par question la liaison avec les performances aux examens partiels ?



<sup>13</sup> Des mesures gouvernementales visant à fixer le nombre de médecins praticiens imposent à la Faculté de Médecine une sélection des étudiants en fin de 3<sup>ème</sup> candidature dès l'an 2000. Un des tests de sélection, réalisé en collaboration<sup>13</sup> avec le SMART, porte sur la Compréhension d'un Texte Scientifique (test CTS). Ce test a été soumis pour la première fois aux étudiants de première candidature en médecine en mai 1998.

Nous avons corrélé les scores des étudiants ( $n = 198$ ) obtenus à chacune des questions du test CTS avec la moyenne de leurs résultats aux partiels.

Le schéma ci-avant permet de visualiser la position des 26 questions (chaque rond) en fonction de la corrélation obtenue. La ligne graduée verticale présente le continuum des valeurs possibles de 1 à -1.

A gauche nous avons positionné les 9 QCM habituelles, celles où la réponse correcte figure parmi les propositions qui ne sont pas SGI. A droite se trouvent les 17 QCM où la réponse correcte est une SGI.

Nous remarquons que les résultats obtenus aux QCM avec réponse correcte SGI ont tendance à être un peu mieux corrélés avec la moyenne des scores aux examens partiels, mais cette différence n'est pas très marquée sur le schéma.

Voici les tableaux récapitulatifs des corrélations récoltées par chaque question avec les partiels. La moyenne de toutes les corrélations des QCM « SGI » est peu élevée (0,22 avec un écart type égal à 0,1) mais presque du double de celle de toutes les QCM « habituelles » (0,13 avec un écart type égal à 0,13). La validité concomitante est donc faible entre les performances des étudiants aux partiels et leurs scores aux QCM, et, encore plus faible quand la réponse correcte n'est pas une SGI.

*Corrélations des scores aux 17 QCM « SGI » avec la moyenne des scores aux examens partiels ( $n = 198$ , corrélations significatives à  $p < 0,05$  marquées en gras)*

aucune						toutes						manque		absurdité		
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
<b>0,3</b>	<b>0,4</b>	<b>0,3</b>	0,1	<b>0,1</b>	<b>0,1</b>	<b>0,2</b>	0,1	<b>0,2</b>	0,1	<b>0,2</b>	0,0	<b>0,3</b>	<b>0,2</b>	<b>0,2</b>	<b>0,1</b>	<b>0,3</b>
<b>0</b>	<b>2</b>	<b>6</b>	2	<b>7</b>	<b>9</b>	<b>5</b>	2	<b>2</b>	2	<b>7</b>	0	<b>1</b>	<b>1</b>	<b>5</b>	<b>6</b>	<b>0</b>

*Corrélations des scores aux 9 QCM « habituelles » avec la moyenne des scores aux examens partiels ( $n = 198$ , corrélations significatives à  $p < 0,05$  marquées en gras)*

1	2	3	4	5	6	7	8	9
-0,04	0,08	<b>0,27</b>	<b>0,15</b>	<b>0,24</b>	-0,02	<b>0,29</b>	-0,02	<b>0,24</b>

### *Les pourcentages de certitude*

A l'aide des pourcentages de certitude l'examineur peut évaluer de façon objective et systématique la capacité des étudiants à auto-évaluer leurs compétences (De Finetti, 1956; Shufford & al., 1966; Van Naerssen, 1962; Leclercq, 1975, 1982, 1993; Hunt, 1977; Bruno, 1986). En demandant aux étudiants de porter systématiquement un jugement sur leurs propres compétences (en ce qui concerne la matière enseignée) à chaque question d'une épreuve, nous nous situons au niveau (6) de la taxonomie de Bloom habituellement très peu évalué. Nous reviendrons plus loin sur les enjeux et les techniques du recueil des pourcentages de certitude. Nous rappellerons ici les conditions méthodologiques qui doivent être rencontrées pour garantir un recueil sans biais des données liées à l'auto-estimation de ses compétences et que Shufford & al. (1966) appellent "Admissible probability measurement procedures". Ces procédures sont mises en œuvre dans le cadre du SMART.

### Lorsque des tarifs sont utilisés ils doivent être conformes à la théorie des décisions

Il s'agit de gratifier une réponse correcte accompagnée d'un degré de certitude élevé d'un meilleur score que si elle était accompagnée d'une certitude faible et inversement pour les réponses incorrectes. Les tarifs du barème des points doivent être calculés de manière à favoriser une seule stratégie : celle qui consiste à dire la vérité. Le barème des points figurant dans le tableau ci-après garantit que l'expression de son intime conviction rapporte plus de points que tout autre stratégie.

### La consigne doit être "probabiliste".

Demander à l'étudiant d'indiquer sa certitude par des termes vagues du type "peu sûr", "moyennement sûr", "très sûr", etc. est à proscrire car ces expressions recouvrent des réalités différentes en fonction des sujets. De plus, avec des termes aussi flous la variabilité est telle chez un même étudiant qu'on ne peut même pas recourir à des traitements ordinaux intra-sujets.

Voici la consigne mise au point par Leclercq (1983, 1993, 1998) et qui est proposée aux enseignants qui souhaitent avoir recours aux pourcentages de certitude dans le cadre des évaluations standardisées réalisées avec l'aide du SMART.

Si vous considérez que votre réponse a une probabilité d'être correcte comprise entre	Ecrivez	Vous obtiendrez les points suivants en cas de	
		réponse correcte (RC)	réponse incorrecte (RI)
0 % et 25 %	0	+ 13	+ 4
25 % et 50 %	1	+ 16	+ 3
50 % et 70 %	2	+ 17	+ 2
70 % et 85 %	3	+ 18	+ 0
85 % et 95 %	4	+ 19	- 6
95 % et 100 %	5	+ 20	- 20

### Des zones correspondant à la précision humaine

Les coupures sur l'axe ne sont pas équidistantes ce qui permet une expression du degré de certitude plus nuancée à l'extrémité supérieure de l'échelle. Ainsi, l'étudiant peut faire la distinction entre 90 % (valeur centrale de la certitude 4) et 97,5 % (valeur centrale de la certitude 5) bien que la différence soit de 7,5 % seulement. Dans le premier cas (90 %) il n'a qu'1 chance sur 10 (1/10) de se tromper tandis que dans le second (97,5 %) il n'a qu'1 chance sur 40 (1/40), soit 4 fois moins. Etablir la même différence au milieu de l'échelle, par exemple entre 40 % (1/1,7) et 47,5 % (1/1,9), n'est pas pertinent car nous ne sommes pas capable de distinguer ces deux derniers « rapports »...

### Le calcul d'indices métacognitifs doit être possible

La consigne utilisée autorise le calcul d'un indice de réalisme basé sur les différences entre les taux d'exactitude et les valeurs centrales des intervalles de probabilité ainsi que le calcul d'un indice de centration basé sur la différence entre la certitude moyenne et le taux d'exactitude moyen et dont le signe détermine la surestimation (+) ou la sous-estimation (-). Nous reviendrons plus loin en détail sur les méthodes et formules de calcul du réalisme (pp. **Erreur ! Signet non défini.**, **Erreur ! Signet non défini.** et **Erreur ! Signet non défini.**) et de la centration (p. **Erreur ! Signet non défini.**).

### *Le questionnement à livres ouverts*

L'examineur qui souhaite mesurer le niveau (2) compréhension et non le niveau (1) de connaissance (ou la reconnaissance parmi une série de propositions dans le cas des QCM) devrait

permettre aux étudiants de répondre aux questions de compréhension en ayant les ressources documentaires à disposition. Leclercq (1993, p. 18) souligne la proximité de cette situation d'évaluation par rapport à la vie courante : « *Open-Book work is the most common situation in adult work, as well as in adult learning. Often, we discover new things to understand and memorise from written text. We read them, we try them (on the computer, the engine or the objects dealing with the content) and, only if we do not understand, do we ask others* ».

### b) La fidélité

Dans quelle mesure un correcteur peut-il prétendre qu'un travail corrigé et classé dans la catégorie « excellent » bénéficierait de la même mention s'il était corrigé dans d'autres conditions (autres correcteurs ou quelques mois plus tard) ? Les problématiques liées à la subjectivité de la correction ont été l'objet d'un courant de docimologie dite « négative » ou « critique » mené par Pieron (1963). Ce dernier et d'autres chercheurs à sa suite ont relevé une série de biais d'évaluation tels que l'inconstance d'un même évaluateur<sup>14</sup> et la discordance entre évaluateurs<sup>15</sup>.

La note (d'une épreuve corrigée) classée dans une catégorie donnée doit l'être de la même façon si elle est traitée par d'autres correcteurs (concordance interjuges) ou/et à un autre moment (consistance intrajuge).

Des procédures d'évaluation automatisées visent à garantir la fidélité des mesures. Les QCM permettent d'échapper à la subjectivité des correcteurs et contribuent ainsi à augmenter la fidélité des évaluations. De plus, la simplicité de correction autorise un traitement informatisé rapide grâce à la lecture optique de marques<sup>16</sup> ou au questionnement interactif (les questions sont alors posées via l'écran d'un ordinateur ce qui permet notamment un feedback après chaque réponse fournie ou/et immédiatement en fin de test). Une autre source de fluctuations est liée à l'impossibilité habituelle d'exprimer son doute ou la sûreté de ses connaissances et nous avons vu qu'il existe des solutions pour permettre un recours systématique aux pourcentages de certitude dans le cadre d'épreuves standardisées à condition de respecter une série de règles méthodologiques.

Les examinés ont des choses à dire sur les évaluations qu'ils subissent, il nous semble important de récolter leurs avis notamment pour augmenter la fidélité des examens. La procédure des feuilles de justification permet à l'étudiant d'exprimer de façon brève (un cadre de réponse limite chaque commentaire à une cinquantaine de mots) son trouble face à certaines questions. Ces justifications subissent ensuite un tri : seuls sont lus les commentaires concernant les réponses incorrectes (afin d'alléger le travail de correction), ceci, après l'analyse des coefficients *rpbis*, amène souvent le professeur à changer les critères de correction d'une évaluation, par exemple, un distracteur s'avère correct et on accorde le point soit à tous les étudiants qui l'ont désigné comme étant exact, soit seulement à ceux qui ont accompagné cette réponse d'un commentaire justifiant

<sup>14</sup> Par exemple l'effet d'ancrage observé par Bonniol (1972) lors de la correction de travaux de valeur moyenne parmi lesquels il introduit des ancrés (copies de valeur soit excellente, soit médiocre) et qui provoquent un effet de contraste sur les travaux suivants.

<sup>15</sup> Cette dernière est illustrée, entre autres, par Agazzi (1967) qui observe à l'occasion de la correction des copies d'un baccalauréat par 6 correcteurs, 70 % des compositions françaises qui sont tantôt admises par les uns et tantôt refusées par les autres. Pieron & al. (1962) estiment qu'il faudrait 16 correcteurs pour stabiliser les notes en physique, 78 correcteurs en composition française et 127 en dissertation philosophique...

<sup>16</sup> Les étudiants répondent en cochant leurs réponses sur des feuilles spéciales qui sont ensuite lues par un dispositif de lecture optique qui peut traiter jusqu'à 6400 copies à l'heure.

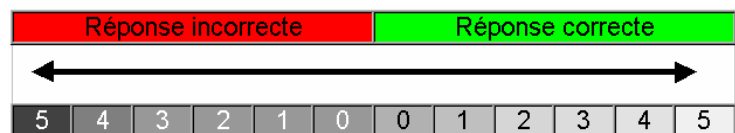
leur choix. Il arrive également qu'une question mal formulée et décelée par cette méthode soit éliminée, on améliore ainsi la fidélité de l'épreuve.

### c) La sensibilité

La mesure doit être précise, refléter des phénomènes subtils.

Habituellement, on ne permet pas aux étudiants d'exprimer leur doute. Dans le contexte des épreuves standardisées ayant recours aux questions fermées (QCM classiques), on se situe alors dans un système de codage binaire des réponses. Soit une réponse est correcte, soit elle est incorrecte (1<sup>ère</sup> ligne de la figure).

Lorsque les évalués peuvent exprimer leur doute à l'aide des pourcentages de certitude, la situation devient plus nuancée. En effet, nous avons vu plus haut que la consigne utilisée dans le cadre des épreuves organisées à l'aide du SMART autorisait l'emploi de 6 degrés de certitude (de 0 à 5). Dès lors, une réponse incorrecte accompagnée d'un degré de certitude maximum (« 5 ») présente la pire des situations, celle où l'étudiant donne une réponse incorrecte en estimant qu'elle a un maximum de chances d'être correcte (à gauche sur le spectre de la figure), on parlera alors de méconnaissance ignorée. A l'opposé (à droite sur le spectre grisé de la figure), on trouve la situation de l'étudiant qui répond correctement avec une certitude maximale, dans ce cas il fait preuve d'une connaissance parfaite. Entre ces deux extrêmes s'ouvre tout l'espace de l'analyse spectrale.



Jans et Leclercq (1999, pp. 308-309) ont étudié en détail l'analyse spectrale appliquée aux performances des étudiants et présentent le caractère éduométrique de l'analyse spectrale : « Nos propositions sont, dans notre esprit, éduométriques par sept aspects :

- Elles s'attachent à rendre compte de la subtilité des phénomènes mesurés. Ainsi la performance n'est plus affaire de tout (1) ou rien (0) : entre ces deux extrêmes, la connaissance partielle et ses divers degrés est désormais pleinement reconnue. Les spectres de performances sont l'illustration de cette approche.
- Elles visent à créer, pour les étudiants des indices individuels indépendants du groupe des étudiants ... « group free scores » (...).
- Elles visent à apprécier des modifications de situation, d'où le besoin de termes tels que « analyse bi-statique, analyse dynamique ... » (...).
- Elles lient l'approche standardisée à l'approche clinique dans le but de comprendre au moins autant que de mesurer (...).
- Elles sont conçues pour être communiquées aux intéressés, pour faire comprendre, un peu comme une psychanalyse « didactique » prépare le patient à devenir un jour expert ; ici à être un expert métacognitif transversal (Jans et Leclercq, 1997) et pas seulement un meilleur maître du contenu spécifique. (...).
- Elles assument la complexité et la multidimensionnalité des compétences et performances par le dégagement de profils de performances et/ou d'apprentissage, par le regroupement des ces profils dans des typologies.
- Elles ont une perspective interventionniste, à savoir le changement (l'amélioration) du profil des performances. ».

En améliorant la sensibilité des instruments de mesure des acquis des étudiants, on augmente les possibilités de différenciation des étudiants grâce à l'affinement de l'échelle des niveaux de maîtrise. En effet, en augmentant la sensibilité de l'instrument par l'emploi des pourcentages de certitude, on peut par exemple distinguer entre (de la situation la plus catastrophique à l'idéal) : (1) méconnaissance erronée (réponse incorrecte et certitude élevée), (2) confusion (réponse incorrecte et certitude moyenne), (3) méconnaissance reconnue (réponse incorrecte et certitude faible), (4) ignorance (réponse correcte et certitude zéro), (5) connaissance douteuse (réponse correcte et certitude faible), (6) connaissance partielle (réponse correcte et certitude moyenne) et (7) connaissance parfaite (réponse correcte et certitude élevée) (Jans & Leclercq 1999, in Depover & Noël, Eds., p. 307). On améliore aussi potentiellement la mesure des progressions individuelles de chaque individu.

#### d) La diagnosticité

L'idée est ici de permettre le diagnostic précis des difficultés d'apprentissage, des processus maîtrisés et de ceux qui ne le sont pas.

Comment diagnostiquer avec précision les difficultés d'apprentissage, les processus maîtrisés et ceux qui ne le sont pas ? Habituellement pour développer des procédures de diagnostic et de remédiation l'examineur recourt à des échelles d'évaluation descriptives. Par exemple, il pose une question ouverte et il la corrige en fonction de  $x$  critères opérationnalisés. La tâche qui est ainsi demandée à l'évaluateur est très complexe, vu l'effort d'analyse, nécessaire pour séparer les différentes catégories de réponses. En outre, il n'est pas possible de procéder à un diagnostic univoque à l'aide d'UNE SEULE question. Il est bien connu, en effet, qu'on ne peut dire d'une question qu'elle mesure A COUP SUR la connaissance, ou la compréhension, ou l'analyse car CELA DEPEND de ce que l'étudiant maîtrise par ailleurs.

VANDEVELDE (1971) signale à ce sujet : « Sans doute 'ce qui se conçoit bien s'énonce-t-il clairement', encore ne peut-on pas confondre comprendre et savoir. L'école se livre à des extrapolations abusives au travers de la connaissance. La connaissance étant la condition sine qua non de la réussite, elle constitue un écran à l'égard de ce que nous voulons mesurer. L'individu examiné peut ne pas posséder la matière sur laquelle porte l'épreuve, n'avoir pas dépensé l'énergie nécessaire à sa mémorisation. Ceci peut bien entendu lui être reproché, mais le grief ne porte dès lors plus sur son inaptitude à comprendre. Nous nous trouvons dans l'impossibilité de conclure à cet égard. (...) C'est pour avoir trop souvent réuni en une même technique d'investigation une gamme importante de facteurs très différents (connaissance, compréhension, application, analyse, etc.) que l'image que nous nous faisons de l'élève est si imprécise. Nos examens, nos épreuves de contrôle scolaire gagneraient beaucoup à se présenter de manière analytique et davantage sur le plan des activités mentales que sur le plan des matières. La question à choix multiple est susceptible de fournir une information précieuse dans cette perspective. »

Le « recouplement » de plusieurs réponses à plusieurs questions permet lui, un diagnostic, comme cela se passe en médecine : la température à elle seule ne permet pas d'identifier la maladie, mais COMBINÉE avec d'autres observations, elle le permet. C'est le principe qui a présidé à la conception de « double check » (Leclercq, 1993), une procédure d'évaluation interactive, qui consiste à poser une question en deux volets *prim* et *bis* et que nous avons présentée plus haut dans le cadre de la validation en référence à un modèle théorique (p. 33).

Ceci dit, même dans les épreuves non interactives il est possible d'améliorer le diagnostic des compétences des étudiants si pour chaque question les catégories de processus mental et de rubrique matière qui lui sont associées ont été identifiées. On est alors en mesure de renvoyer vers l'étudiant des feedbacks automatisés plus analytiques du type : « *Votre taux de réussite en ce qui concerne le chapitre I est de 45%. Vous avez répondu correctement à 9 questions sur 20 dont 8 questions sur 10 de connaissance et 1 question sur 10 d'application...* ». Cet affinement du diagnostic peut aussi aboutir à des prescriptions du type : « *Etant donné ces taux de réussite revoyez les exercices proposés en fin du chapitre I...* ».

L'utilisation des pourcentages de certitude permet aussi de proposer un diagnostic méta-cognitif. Nous verrons plus loin qu'on peut calculer le réalisme des au départ des informations qu'ils nous fournissent en accompagnant chacune de leurs réponses d'un pourcentage de certitude.

### e) La praticabilité

Garantir l'objectivité des corrections, l'évaluation systématique des processus mentaux en jeu ainsi que l'augmentation de la sensibilité et de la « diagnosticité » impliquent dans le cadre d'examens classiques<sup>17</sup> que l'on y consacre de l'énergie et du temps. Il est possible de poser une question ouverte à réponse écrite longue notée selon  $x$  points de vue différents à l'aide de  $y$  critères préalablement opérationnalisés pour chacun, et il en résulte une correction complexe. Or, comme dans la plupart des universités européennes (Gibbs, Jenkins & al., 1992), on assiste aujourd'hui à une explosion du nombre d'inscriptions d'étudiants à la FAPSE-ULg<sup>18</sup>, et, si on multiplie par le nombre d'étudiants le temps passé à corriger un examen dans ces conditions on en arrive rapidement à un constat d'impraticabilité.

Le SMART propose des techniques d'évaluation informatisées, notamment par QCM-SGI avec degrés de certitude et lecture optique de marques ainsi qu'un soutien méthodologique.

De plus en plus d'enseignants de l'Université de Liège et de Hautes Ecoles font appel au SMART. Dans plusieurs examens, deux types de questions sont proposés en symbiose. D'une part, un grand nombre de QCM-SGI (en général une trentaine) avec degrés de certitudes, offrent une série d'avantages décrits par Leclercq (1986) : représentativité de l'échantillon des questions, simplicité de correction, objectivité de la correction, possibilité d'évaluer systématiquement les niveaux de compréhension, d'analyse et d'évaluation/jugement. D'autre part, un petit nombre de questions ouvertes (en général une ou deux) permettent d'évaluer l'esprit de synthèse, l'originalité, la créativité, la capacité à organiser une réponse, ce qui n'est pas possible en ayant recours uniquement aux QCM fussent-elles SGI.

### f) L'équité

Tous les étudiants doivent être traités de façon juste, en principe de la même manière. Veiller à ce qu'une épreuve soit équitable c'est notamment poser le problème de la standardisation du test.

<sup>17</sup> Nous entendons par *examens classiques*, les évaluations ayant recours aux questions ouvertes à réponse construite longue posées soit à l'oral, soit à l'écrit.

<sup>18</sup> D'environ 200 en 1986-1987, le nombre d'inscriptions en première candidature de la FAPSE-ULG est passé à 400 en 1994-1995.

Une standardisation parfaite implique :

- que tous les étudiants d'une section soient soumis à la même épreuve (mêmes questions pour tous) ;
- que les conditions d'administration de l'examen soient identiques pour tous les examinés ;
- que les modalités de correction soient uniformisées, que toutes les copies soient corrigées de la même manière.

Dans les faits, une standardisation idéale : même heure, même dispositions matérielles (sièges, éclairage, ...), même degré de familiarité des examinés avec les techniques de questionnement, même consignes, même entraînement, même durée, etc. est difficilement réalisable, voire impossible, mais il importe d'essayer de s'en approcher le plus possible.

### **g) La communicabilité**

Les informations non confidentielles relatives au déroulement du processus SMART doivent être communiquées et comprises par les partenaires engagés dans la réalisation des évaluations.

Cet objectif de « communicabilité » pose notamment le problème des feedbacks vers les étudiants après les épreuves. Les Feedbacks Via Internet (FVI) apportent une solution technologique en permettant la consultation via le web des feuilles de résultats individualisées après les épreuves en amphithéâtre.

### **h) L'authenticité**

Rappelons notre définition de « 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.*

Des QCM comportant des cas à analyser soutenus par des documents, par exemple dans le contexte de la médecine humaine : des protocoles de laboratoire, des radio, des lettres de confrères, ... et mis en situation, par exemple à l'aide de séquences vidéoscopées, sont certainement de nature à rencontrer notre critère d'authenticité. Il s'agit là de modalités de questionnement que nous comptons implémenter dans le projet de plate-forme *e-C&QCST*.

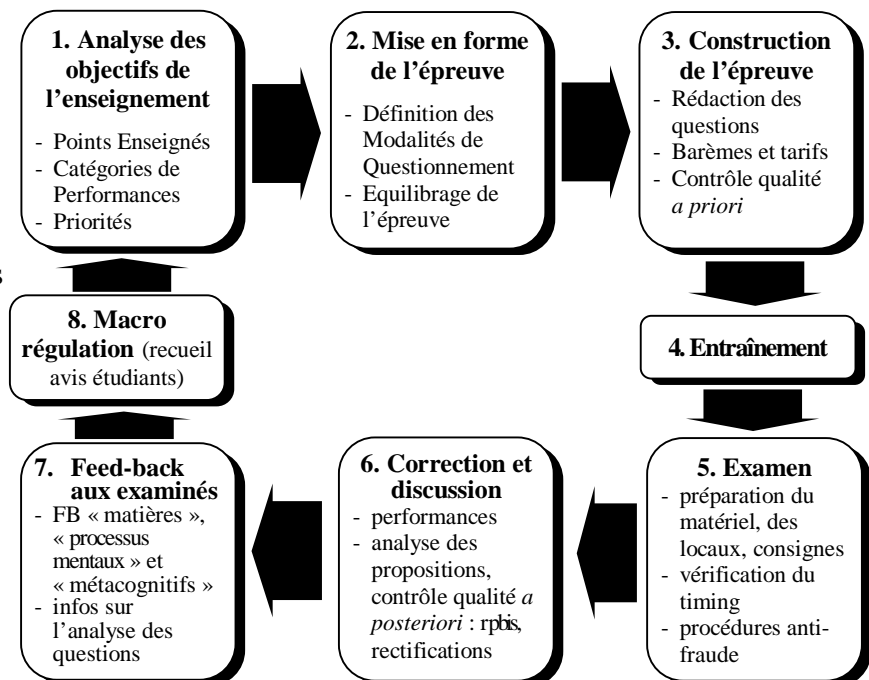
### 3. Cycle de construction en « spirale de qualité » des évaluations standardisées et dispositifs d'ingénierie docimologique associés

Le modèle que nous présentons est à situer dans le contexte de la construction des tests standardisés visant à évaluer les acquis des apprenants. Le modèle est particulièrement efficace lorsqu'il s'agit d'évaluer de grands groupes. Il est proposé dans le cadre de la construction des examens standardisés au sein d'universités ou de Hautes Ecoles ainsi que dans le contexte des évaluations en entreprise (par exemple chez FORTIS BANK) ou au sein d'administrations (par exemple au sein de l'Institut de Formation des Administrations fédérales [IFA]) lorsque les professeurs ou les formateurs souhaitent recourir aux services du Système Méthodologique d'Aide à la Réalisation de Tests (SMART). Il existe bien entendu d'autres modèles, par exemple celui du processus de construction d'un instrument de mesure proposé par Laveault & Gregoire (1997, pp. 79-122). Dans leur livre, ces auteurs proposent 5 étapes pour la réalisation d'un instrument de mesure. Etape 1 : détermination des utilisations prévues du test. Etape 2 : définition de ce que l'on souhaite mesurer. Etape 3 : création des items. Etape 4 : évaluation des items. Etape 5 : détermination des propriétés métriques du test définitif. Ils distinguent aussi des procédures de construction différentes selon que l'on élabore des tests d'acquis scolaires à visées sommatives ou à visées formatives.

Notre modèle en 8 étapes englobe d'autres phases non couvertes par Laveault & Gregoire. Il met notamment l'accent sur l'entraînement des étudiants et sur la fourniture de feedbacks informatifs et rapides à ces derniers. L'idée étant de fournir autant que possible des informations aux examinés en situation d'échec en vue de les aider à atteindre le niveau de compétence exigé lors d'une session d'examen ultérieure. Le modèle qui sera exposé s'insère aussi dans une perspective de travail en « spirale de qualité » qui permet d'améliorer tout le processus de construction des épreuves.

Les huit étapes de la gestion des tests standardisés (Gilles & Leclercq, 1995 ; Gilles, 2002) présenté ci-contre, permettent de décrire un processus cyclique de réalisation des tests standardisés. Elles sont inspirées du schéma des grandes phases de construction d'un examen décrit par De Landsheere (1980, p. 65).

Le respect de ces huit phases de construction et des huit objectifs « qualité » précédemment décrits nous paraît crucial si nous souhaitons réaliser des évaluations standardisées de qualité.



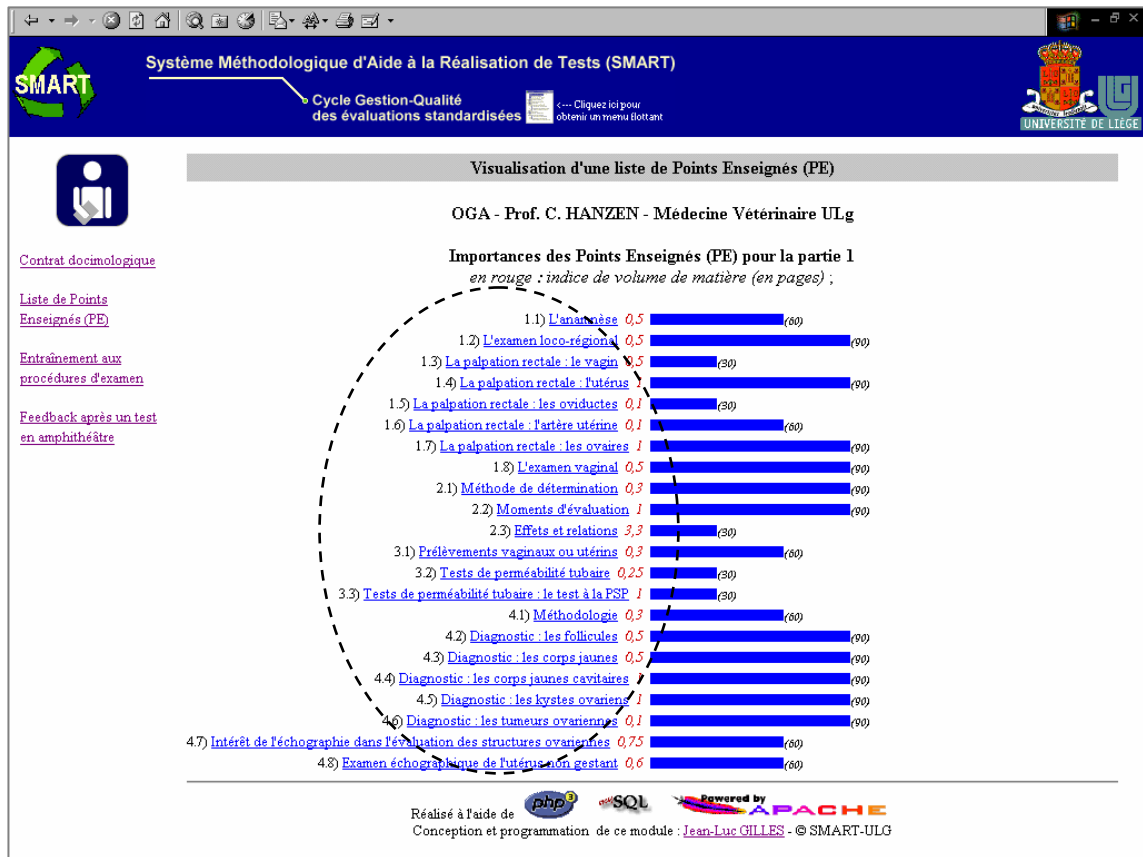
## a) L'analyse des objectifs de l'enseignement ou de la formation

Cette étape a pour but de clarifier et de mettre en évidence les objectifs susceptibles d'être évalués dans le contexte d'une épreuve standardisée universitaire. D'un point de vue pragmatique, l'idée est de proposer une méthode qui permet un inventaire rapide des objectifs de l'enseignement pour lequel un examen standardisé est envisagé. La mise en œuvre de cette méthode sera rendue encore plus aisée si le professeur/formateur prend soin de baliser ses notes de cours par des codes lui permettant de repérer les points qu'il considère comme formant des unités de matière.

La procédure d'analyse des objectifs de l'enseignement que nous proposons dans le contexte du SMART se déroule en quatre phases.

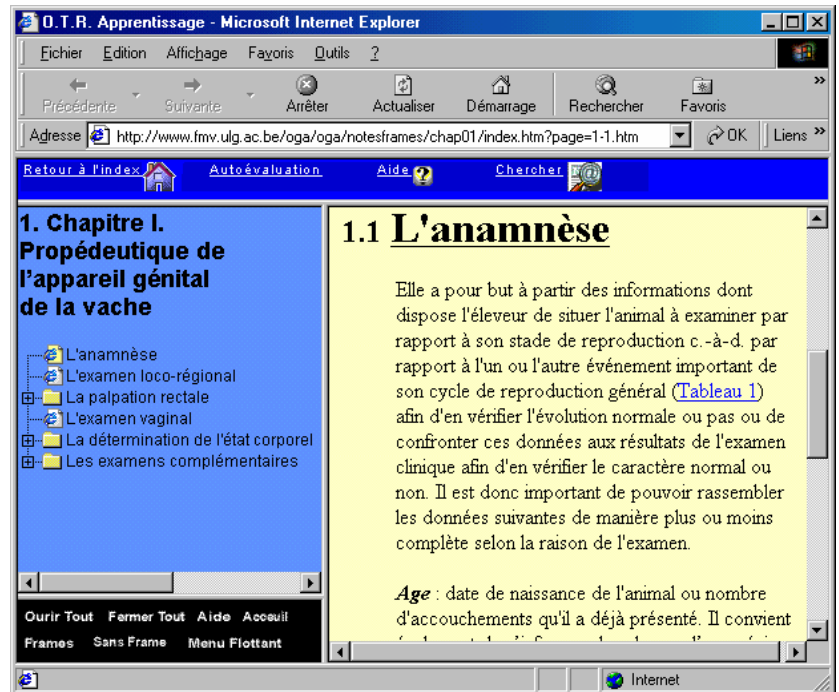
Phase 1 : dans un premier temps l'enseignant liste les Points Enseignés (PE) (ou les rubriques de matière qui ont été abordées au cours) et fixe leur importance.

Voici un exemple de liste de Points Enseignés liés au cours d'Obstétrique des Grands Animaux (OGA) pour lequel nous avons utilisé cette méthode lors de la construction d'épreuves standardisées (Castaigne & Gilles, 2000 ; Castaigne, Gilles & Hansen, 2001). Les traits bleus à droite permettent de visualiser le degré d'importance attribué par le professeur à chacun des points enseignés (0 = importance minimum et 100 importance maximum).



Sur l'écran de l'interface expérimentale<sup>19</sup> ci-dessus par nous avons entouré en pointillés une série de PE sélectionnés dans le cadre du cours OGA.

Grâce au support de l'enseignement à distance offert par le service du Prof. Hansen aux étudiants du cours OGA, des liens entre l'interface du SMART et le site Internet professeur ont pu être créés. En cliquant sur un point enseigné (écran précédent) l'étudiant pouvait obtenir la matière du cours concernée<sup>20</sup> (exemple ci-contre).



Phase 2 : dans un second temps l'enseignant qui procède à la validation du contenu de son épreuve identifie les Catégories de Performances (CP), c'est-à-dire les processus mentaux que les étudiants doivent exercer sur les points enseignés.

Codes PE	Intitulés PE	Connaissance	Compréhension	Application	Analyse	Synthèse	Evaluation
4.1	Méthodologie	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4.2	Diagnostic : les follicules	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4.3	Diagnostic : les corps jaunes	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4.4	Diagnostic : les corps jaunes	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4.5	Diagnostic : les kystes ovariens	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4.6	Diagnostic : les tumeurs ovariennes	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4.7	Intérêt de l'échographie dans	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4.8	Examen échographique de	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

<sup>19</sup> Il s'agit d'une version expérimentale créée par Jean-Luc Gilles (Gilles 2002).

<sup>20</sup> Le contenu du cours du professeur C. Hansen a été porté sur Internet par le Dr J.-L. Castaigne. Nous tenons à les remercier tous deux pour leur précieuse collaboration.

C'est la taxonomie des objectifs cognitifs de Bloom (1969) qui fut choisie dans le cadre de la construction des épreuves du cours OGA. L'identification des CP permet d'établir une liste des binômes [PE x CP] où pour chaque point enseigné l'examineur est amené à associer une ou plusieurs catégories de performances. L'interface ci-dessous développée par le SMART est disponible via l'Internet et permet à l'enseignant d'associer à chaque Point Enseigné (PE) une ou plusieurs catégories de Performances (CP). Dans le cadre de l'exemple ci-dessus, le point enseigné « 4.2 Diagnostic : les follicules » est associé à trois catégories de performances issues de la taxonomie de Bloom : « *Connaissance* » (réactualisation de contenus mémorisés), « *Compréhension* » (interprétation, traduction avec accès aux ressources documentaires) et « *Analyse* » (décomposition, classification, organisation). On obtient ainsi trois binômes [PE x CP]. Ces binômes peuvent ensuite être classés en fonction de l'importance accordée par l'enseignant (établissement des priorités évaluatives).

### b) Mise en forme de l'épreuve

**Phase 3 :** Lors de cette étape l'enseignant ou le formateur qui utilise la procédure de validation du contenu définit les Modalités de Questionnement (MQ) qu'il compte utiliser. Les MQ sont composées d'un type de questionnement et d'une série de caractéristiques associées. Exemples de MQ : le type de questionnement « QCM » accompagné de « Solutions Générales Implicites (SGI) » à « livres ouverts<sup>21</sup> » ou le type de questionnement « QROC<sup>22</sup> » accompagné de la technique des « degrés de certitude », ...

Ces Modalités de Questionnement (MQ) sont ensuite associées aux Catégories de Performances (CP) mises en évidence à l'étape précédente d'analyse de l'enseignement. On obtient alors une deuxième série de binômes intitulés « [CP x MQ] ». Dans le cadre de la méthode de validation du contenu, l'interface proposée par le SMART permet à l'évaluateur de créer ses binômes [CP x MQ] en sélectionnant les éléments constitutifs dans des listes (les informations sont ensuite mémorisées dans une base de données).

[CP x MQ] (4)  MQ :

Les nombreuses possibilités offertes par la combinaison des modalités de questionnement et des catégories de performances ont été décrites par ailleurs dans Leclercq & Gilles (1995).

**Phase 4 :** Enfin, pour chaque Point Enseigné mis en évidence à l'étape de l'analyse des objectifs de l'enseignement (*Phase 1*), le professeur peut concevoir un ou plusieurs trinômes [PE x CP x MQ]. Chaque trinôme est donc formé d'un Point Enseigné, d'une Catégorie de Performance et d'une Modalité de Questionnement. Cela donne par exemple le trinôme : « *Connaissance (réactualisation de contenus mémorisés)* [CP] » x « *liée au diagnostic des follicules* [PE] » x « *à l'aide d'une QCM-SGI avec degrés de certitude* [MQ] ».

Une liste des trinômes peut alors être générée et proposée à l'examineur qui sélectionne en fonction de l'importance qu'il leur accorde les trinômes qui feront l'objet d'une ques-

<sup>21</sup> Les étudiants disposent alors lors de l'examen de ressources documentaires spécifiées par l'examineur, dès lors les questions peuvent réellement mesurer les performances des étudiants liées au niveau taxonomique « compréhension » (selon la taxonomie des objectifs cognitifs de Bloom, 1969).

<sup>22</sup> QROC = Question à Réponse Ouverte Courte.

tion lors de l'épreuve. On aboutit ainsi à un plan d'épreuve qui guidera la création des questions à l'étape suivante « c) Construction de l'épreuve ».

## c) Construction de l'épreuve

On procédera ici à la réalisation des questions en tenant compte de règles méthodologiques de construction lorsqu'elles existent (par exemple les règles de rédaction des QCM de Leclercq, 1986). Voici un plan d'épreuve conçu par le professeur Lion dans le cadre des interrogations hebdomadaires (année académique 1999-2000) proposées aux étudiants ingénieurs qui ont suivi le cours de Physique (Lion, 2000). Dans le contexte de ces épreuves, il n'y avait pas de définition de Modalités de Questionnement (MQ) dans la mesure où toutes les questions étaient de type « QCM » avec « Solutions Générales Implicites » (SGI) et « pourcentages de certitude ».

Un tel plan permet de guider et de rationaliser la réalisation des questions, voici la 1<sup>ère</sup> question de cette épreuve dont le PE est « *Interaction entre deux boucles* » et la CP « *Compréhension* » :

<b>Interro 14B</b>		<b>Nombre de questions: 10</b>					
Pour revenir à l'accueil ou imprimer, cliquez sur "continuer".							
Légende : RC = Réponse Correcte ; NSP = Nombre de solutions proposées ; RM = Rubrique matière ; PM = Processus Mental.							
Code	Titre	Type	RC	NSP	RM	PM	
5 094 451	Couple sur boucle dans B	Problème simple	8	4	5	4	
5 095 466	Interaction entre 2 boucles	Compréhension	2	5	5	2	
5 101 441	Para- et diamagnétisme	Connaissance	4	4	5	1	
5 103 460	Energie d'orientation d'un dipôle	Compréhension	4	4	5	2	
5 112 446	Faraday - Boucle tournant dans B	Problème simple	4	4	5	4	
5 112 452	Unité de B	Connaissance	4	5	5	1	
5 112 453	Bobine dans B variable	Problème simple	1	4	5	4	
5 113 461	Barre sur rails dans B	Problème simple	8	4	5	4	
5 114 443	Loi de Faraday	Compréhension	3	4	5	2	
5 114 450	Unité de la circulation de E	Connaissance	4	4	5	1	

**Q2.** Deux fils conducteurs circulaires se trouvent dans deux plans parallèles; la droite passant par les centres des cercles ainsi constitués est perpendiculaire aux deux plans; si les deux boucles sont parcourues par des courants de même sens, elles vont :

1. se repousser;
2. s'attirer;
3. n'exercer aucune force l'une sur l'autre;
4. s'attirer si les intensités des courants sont égales;
5. se repousser si les intensités des courants sont égales.

Lors de cette étape, un contrôle qualité *a priori* peut avoir lieu en soumettant le questionnaire à la critique de collègues experts du contenu. Des docimologistes peuvent aussi effectuer à la demande des professeurs une relecture « formelle » de leurs QCM suivie d'un entretien où ils discutent des pistes à suivre en vue d'améliorer la qualité des questions.

L'enseignant définira également les barèmes et tarifs liés aux questions. Un « poids » peut aussi être associé à chaque question en fonction de l'importance que lui accorde l'examineur (voir plus haut).

### d) Entraînement des évalués

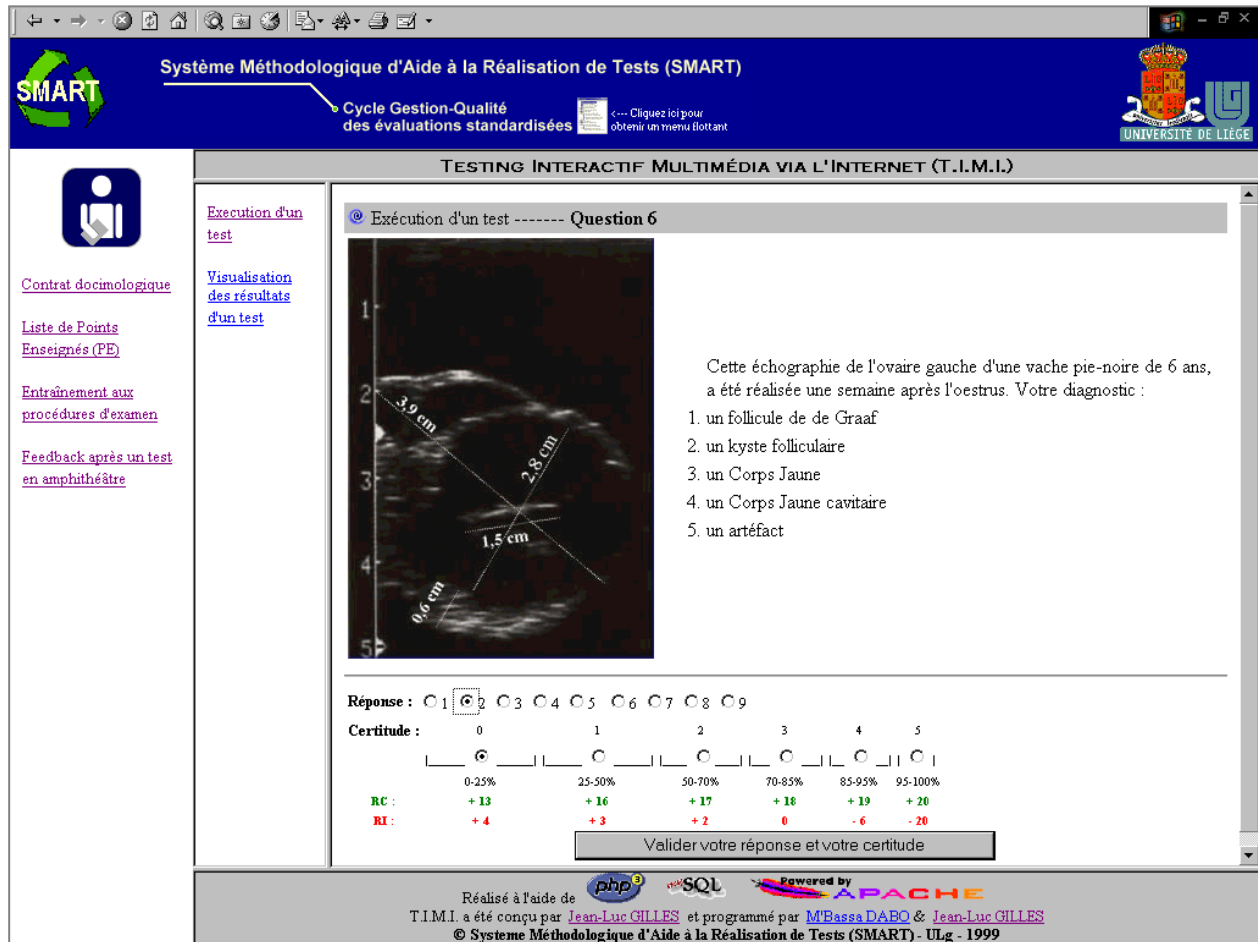
Cette étape est particulièrement importante lorsqu'on utilise des méthodes de questionnement sophistiquées auxquelles les évalués sont peu habitués. Par exemple, il est recommandé d'entraîner les apprenants à l'utilisation des Solutions Générales Implicites et des degrés de certitude avant la première épreuve certificative.

Nous préconisons au moins un test (quiz) d'entraînement en amphithéâtre pour les étudiants qui entrent à l'université et qui seront amenés à utiliser les dispositifs du SMART de Lecture Optique de Marques (LOM). Un quiz en fin de cours permet ainsi de se familiariser avec les formulaires spéciaux prévus pour la lecture optique et les méthodes de questionnement associées.

Pour certains cours, les questions des examens qui ont eu lieu lors des années académiques précédentes (et qui sont donc « brûlées ») sont disponibles via l'Internet. Nous avons intitulé cette forme d'entraînement « *WebQuiz* » (Gilles, 1998b). A l'aide de ces derniers, l'étudiant peut répondre de chez lui ou d'une des salles publiques informatisées de l'ULg. Ces tests peuvent se faire entre deux cours et ne sont pas sanctionnés. L'ordre de réponse aux questions est déterminé par l'étudiant qui peut lire les questions avant de répondre et commencer par celle de son choix.

Récemment, tirant parti d'expériences précédentes (Leclercq & Gilles, 1993 ; Leclercq, Rommes & al., 1998) et plus particulièrement d'une utilisation du multimédia en intranet (Gilles & al., 1999), des dispositifs de Testing Interactifs Multimédias via Internet (*TIMI*) ont été mis au point dans le cadre du SMART. Ces *TIMI* permettent l'utilisation combinée dans les tests de séquences multimédias enregistrées sur CD-ROM (fournis aux étudiants) en complémentarité avec les QCM-SGI stockées dans des bases de données accessibles via l'Internet.

Les *TIMI* permettent aussi aux étudiants de recevoir un feedback immédiat à propos de la réponse correcte attendue après chaque question ainsi qu'un bilan de leurs performances cognitives (taux de réponses correctes, moyenne) et métacognitives (réalisme, tendance à se sur ou sous-estimer).

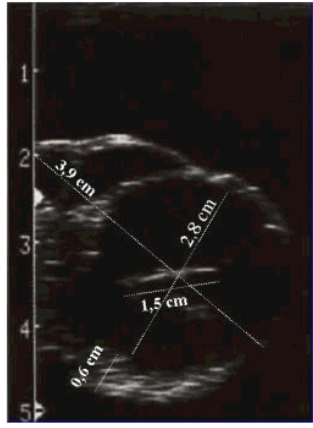


**Systeme Méthodologique d'Aide à la Réalisation de Tests (SMART)**

Cycle Gestion-Qualité des évaluations standardisées

TESTING INTERACTIF MULTIMÉDIA VIA L'INTERNET (T.I.M.I.)

Exécution d'un test ----- Question 6



Cette échographie de l'ovaire gauche d'une vache pie-noire de 6 ans, a été réalisée une semaine après l'oestrus. Votre diagnostic :




1. un follicule de de Graaf
2. un kyste folliculaire
3. un Corps Jaune
4. un Corps Jaune cavitare
5. un artefact

Réponse :  1  2  3  4  5  6  7  8  9

Certitude :

	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	0	1	2	3	4
	0-25%	25-50%	50-70%	70-85%	85-95%
RC :	+13	+16	+17	+18	+19
RI :	+4	+3	+2	0	-6

Valider votre réponse et votre certitude

Réalisé à l'aide de   Powered by 

T.I.M.I. a été conçu par Jean-Luc GILLES et programmé par M'Bassa DABO & Jean-Luc GILLES  
 © Systeme Méthodologique d'Aide à la Réalisation de Tests (SMART) - ULg - 1999

Enfin, il existe aussi un entraînement à l'utilisation des pourcentages de certitude qui est accessible via l'Internet<sup>23</sup> sous la forme d'une activité intitulée GUESS (Leclercq & Gilles, 1994). Lors de cet entraînement les étudiants sont amenés à accompagner de nombreuses réponses de leurs pourcentages de certitude. Un feedback est fourni en direct tout au long de l'exercice à chaque étudiant qui peut voir l'évolution de son réalisme, de sa tendance à se sur ou sous-estimer.

L'entraînement des évalués est aussi pris au sérieux ailleurs dans le monde. Par exemple le Chicago Board of Education (1999) édite pour les enseignants des « Chicago Public Schools » (USA) un manuel d'instruction en vue de préparer leurs étudiants à la passation des tests standardisés. L'ouvrage s'intitule « *Preparing Your High School Students to Take Standardized Tests* » et est accessible via l'Internet<sup>24</sup>, les auteurs y exposent une « *test-taking skills and abilities checklist* » de neuf conseils à répercuter auprès des étudiants ou à effectuer avec eux : « (1) *Follow the directions closely*, (2) *Budget time appropriately*, (3) *Check work carefully*, (4) *Read the entire item and all answers*, (5) *Answer easier questions first and persist to the end*, (6) *Make educated guesses*, (7)

<sup>23</sup> Adresse Internet : <http://www.ulg.ac.be/cafeim/guess/guessdc6.htm>. La version Internet de GUESS a été conçue par D. Leclercq & J.-L. Gilles et programmée dans le langage JAVA par M. Hurard.

<sup>24</sup> Adresse Internet : <http://intranet.cps.k12.il.us/Assessments/Preparation/preparation.html>

*Use test item formats for practice, (8) Review the practice items and answers with students, (9) Practice using answers sheets ».*

Enfin, signalons que depuis l'année académique 1998-1999, le SMART met aussi à la disposition des enseignants et des formateurs un système de boîtiers de vote électronique qui permet aussi d'entraîner les étudiants en amphithéâtre (Leclercq & al., 1999 ; Gilles & al., 2000). L'équipement comprend une centaine de boîtiers, une station de réception, un ordinateur et un vidéo projecteur portables.



Les boîtiers de vote ressemblent à des télécommandes sans fil à 10 chiffres (0 à 9) permettant à 100 étudiants (ou à 100 groupes d'étudiants) réunis dans un même amphithéâtre de fournir une réponse codée par un chiffre (donc à des QCM). Ces réponses sont captées par un "concentrateur" et traitées par un ordinateur portable contenant les fichiers de questions (rédigées au préalable par le professeur). Un logiciel (*INTERACT*<sup>25</sup>) qui permet d'afficher immédiatement sur l'écran du professeur les résultats d'ensemble. Ces résultats sont mémorisés dans un fichier.

Le professeur peut, s'il le veut, projeter à l'aide du vidéo projecteur son propre écran pour que les étudiants eux aussi aient, en direct, l'image de l'ensemble des réponses.

#### e) Mise en œuvre de l'épreuve

L'épreuve proprement dite nécessite une série de préparatifs. La standardisation d'une épreuve implique une définition précise des modalités d'utilisation du test, des consignes. Le SMART apporte aussi son expertise dans ce domaine.

L'examineur veillera lors de l'épreuve à mettre en place des procédures anti-fraude qu'il aurait tort de négliger.

#### *Les examens ayant recours à la Lecture Optique de Marques (LOM)*

Plusieurs modalités de questionnement sont disponibles lorsque la lecture optique de marque est utilisée dans le cadre de l'aide proposée par le SMART :

- Les QROC ou Questions à Réponses Ouvertes Courtes : l'étudiant répond par un mot ou une locution lu(e) par un correcteur humain qui marquera sa note dans la case *ad hoc* (correct / incorrect) sur le *formulom* de l'étudiant, avant traitement de la feuille.
- Les QCL ou Questions à Choix Large : l'étudiant a le choix entre des centaines de réponses possibles (999 maximum), ce qui constitue une procédure intermédiaire entre les QCM et les QRO (Réponses Ouvertes) car l'étudiant doit d'abord penser (se rappeler, évoquer) la solution

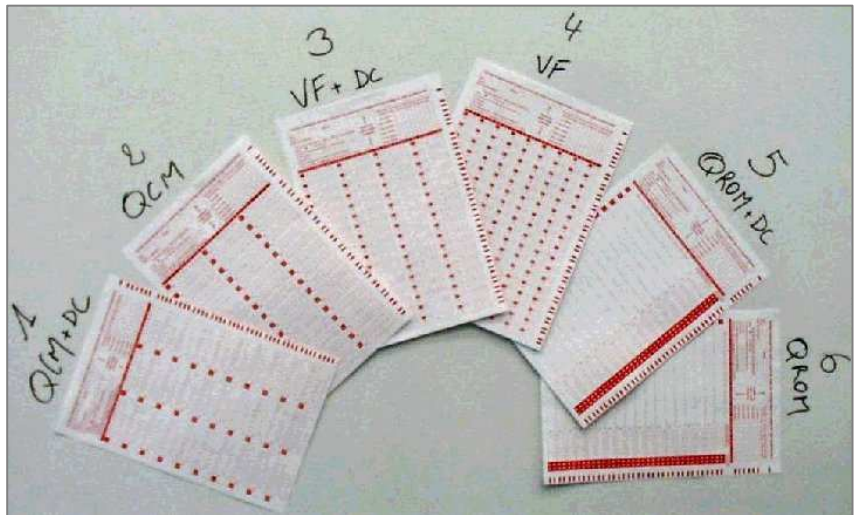
<sup>25</sup> *INTERACT* est un logiciel produit par la firme Soft Concept.

correcte AVANT d'aller voir dans la liste (index sous forme de mini dictionnaire) comment la coder.

- Les QCM SGI ou à Solutions Générales Implicites : où en plus des (3 à 5) solutions dactylographiées, l'étudiant doit considérer comme possible les 4 SGI que sont « 6. Aucune », « 7. Toutes », « 8. Manque de données » et « 9 Absurdité dans l'énoncé ». Cette consigne convient particulièrement aux interrogations à livres ouverts, car elle permet de tester spécialement la compréhension et la vigilance cognitive.
- Les QVF ou Questions Vrai Faux
- Les degrés de certitude peuvent s'ajouter à toutes les formes de questionnement proposées ci-dessus et permettent de calculer outre les résultats cognitifs classiques (taux d'exactitude), un indice métacognitif de réalisme (sur ou sous-estimation) pour chaque étudiant.

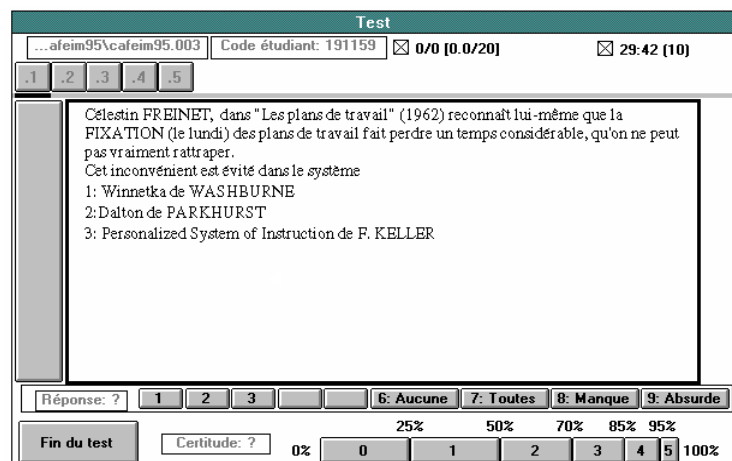
Voici une vue des différents *formuloms*<sup>26</sup> développés par le SMART pour permettre un traitement automatisé des modalités de questionnements qui viennent d'être présentés.

Lors de certaines épreuves, les évaluateurs autorisent de brèves justifications des réponses sur des feuilles *ad hoc*. Il est en outre convenu que le(s) correcteur(s) ne lira(ont) que les commentaires concernant les réponses incorrectes. Avec ce système, la justification ne peut donc QUE bénéficier à l'évalué.



### Les examens ayant recours au testing interactif en intranet (WINCHECK)

WINCHECK (Leclercq & al., 1994) est un logiciel d'évaluation interactive développé par le Service de Technologie de l'Education de l'Université de Liège (ULg). Le programme fonctionne dans l'environnement Windows et permet l'utilisation du « double check ». WINCHECK est disponible via les ordinateurs et le réseau Intranet du Centre d'Auto-Formation et d'Evaluation Interactives Multimédias (CAFEIM) de la FAPSE-ULg. L'évaluateur peut se constituer des banques de questions (QCM-SGI) et créer



Test

...afeim95vcafeim95.003 Code étudiant: 191159  0/0 [0.0/20]  29:42 (10)

1 2 3 4 5

Célestin FREINET, dans "Les plans de travail" (1962) reconnaît lui-même que la FIXATION (le lundi) des plans de travail fait perdre un temps considérable, qu'on ne peut pas vraiment rattraper.  
 Cet inconvénient est évité dans le système

1: Winnetka de WASHBURNE  
 2: Dalton de PARKHURST  
 3: Personalized System of Instruction de F. KELLER

Réponse: ? 1 2 3 6: Aucune 7: Toutes 8: Manque 9: Absurde

Fin du test Certitude: ? 0% 0 1 2 3 4 5 100%

<sup>26</sup> *Formulom* est le terme technique utilisé par le SMART pour désigner le formulaire de réponses spécial destiné à la lecture optique de marques (ils sont imprimés en encre aveugle rouge sur du papier spécial équerre).

des tests en sélectionnant les questions en fonction de critères liés aux contenus, aux processus mentaux et aux objectifs.

Avant le premier examen ayant recours à ce logiciel, tous les étudiants du premier cycle sont invités à s'entraîner (exercice sans sanction) à l'utilisation du programme *WINCHECK* et de la procédure « A livre ouverts ». Le test d'entraînement est composé d'une dizaine de questions portant sur un magazine.

A la FAPSE-ULg les étudiants prennent rendez-vous pour l'examen en réservant un ordinateur au CAFEIM dans une plage horaire qui leur convient.

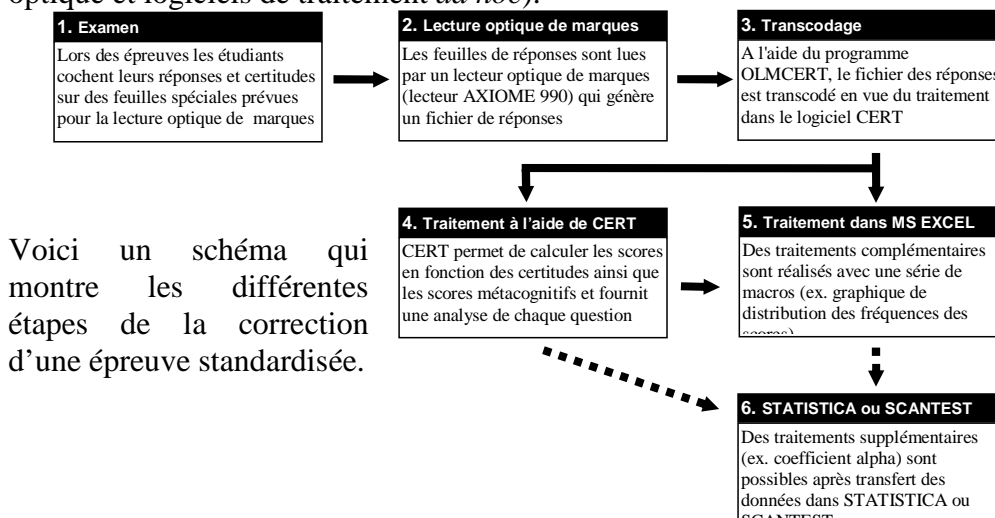
Enfin, signalons aussi l'importance de recueillir des informations sur la qualité des questions lors de l'examen proprement dit. Les étudiants n'hésitent pas à signaler les problèmes qu'ils perçoivent dans les questions lors des épreuves. Collecter leurs avis est utile dans la mesure où ils apportent un éclairage qualitatif qui permet de mieux interpréter les indices de qualité des questions calculés lors de l'étape suivante de correction et discussion.

### f) Correction et discussion des résultats

Durant l'étape de correction une série de contrôles de qualité ont lieu lors de la lecture informatisée des réponses fournies par les étudiants. En ce qui concerne la qualité des questions, des contrôles de qualité *a posteriori* sont effectués à l'aide d'indices de discrimination tels que la corrélation bisériale de point classique (*rpbis classique*) calculée pour chaque proposition au sein des QCM.

Lors de cette étape, le SMART communique les résultats de l'analyse de la qualité des questions à l'examineur. Lorsque la situation l'exige, cet envoi d'information peut être suivi d'une discussion entre ce dernier et un chercheur du SMART à propos des mesures correctives (suppression de questions, valorisation de propositions au départ considérées comme incorrectes, ...) à mettre en place pour rectifier l'épreuve.

Le SMART est équipé d'une chaîne complète de lecture optique de marques performante et fiable pour produire des questionnaires, les lire et les traiter (duplicatrice digitale, lecteur optique et logiciels de traitement *ad hoc*).



Après l'examen, la lecture optique des réponses des étudiants et le transcodage du fichier au format du logiciel de traitement CERT<sup>27</sup> (Boxus & al., 1991), l'analyse de la qualité des questions peut réellement débiter.

### Le traitement à l'aide du programme CERT

Les formuloms sont lus par un lecteur optique de marques qui génère un fichier informatisé contenant les réponses des étudiants. Ce fichier est ensuite converti au format du logiciel CERT utilisé

pour la correction. CERT fournit l'analyse de chaque QCM de l'épreuve en trois lignes par question.

Pour chaque solution proposée (de 1 à 9 pour ces QCM-SGI) sont en effet fournis (0 = omission) dans ces trois lignes respectivement :

Processus : Pas de sélection.		Matière : Pas de sélection.									
SOL: 0		1	2	3	4	5	6	7	8	9	
(1)	Q 26	0.00	26.40	3.37	7.30	33.71	5.62	11.24	1.69	1.12	9.55
(2)	Rbis	0.00	0.01	0.11	-0.27	0.25	-0.03	-0.03	-0.04	-0.12	-0.14
(3)	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	44.09	44.30	36.50	12.50
	Q 28	0.00	1.12	2.81	44.94	26.97	10.67	8.99	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

- (1) le pourcentage d'étudiants qui ont choisi la proposition ;
- (2) le coefficient *rpbis classique* ;
- (3) la certitude moyenne. L'écran présenté ci-dessus montre comment se présente l'écran ou le listing (les statistiques concernant la réponse correcte sont encadrées de deux barres verticales).

Sans entrer dans les détails ici, disons que le *rpbis classique* est un indice de discrimination qui varie dans l'intervalle [-1, 1] qui permet d'évaluer dans quelle mesure une proposition discrimine les étudiants « forts » des étudiants « faibles » (du point de vue du score total obtenu au test). L'analyse d'une épreuve à l'aide de ce dispositif peut amener l'évaluateur à supprimer une question, par exemple lorsque la réponse correcte récolte un *rpbis classique* négatif ou proche de zéro et certains distracteurs des *rpbis classiques* positifs marqués. Dans l'exemple ci-dessus, la question 27 devrait être analysée car le *rpbis classique* de la réponse correcte (SOL 3) est faible (0,09) et en particulier celui du premier distracteur (SOL 1) presque aussi élevé (0,07).

Il arrive que l'évaluateur décide d'accepter le choix d'un distracteur comme réponse correcte au vu des *rpbis classiques* et des éventuels commentaires de justification des étudiants. Il est également possible d'exporter les résultats des évaluations dans les logiciels EXCEL et STATISTICA en vue de traitements plus poussés.

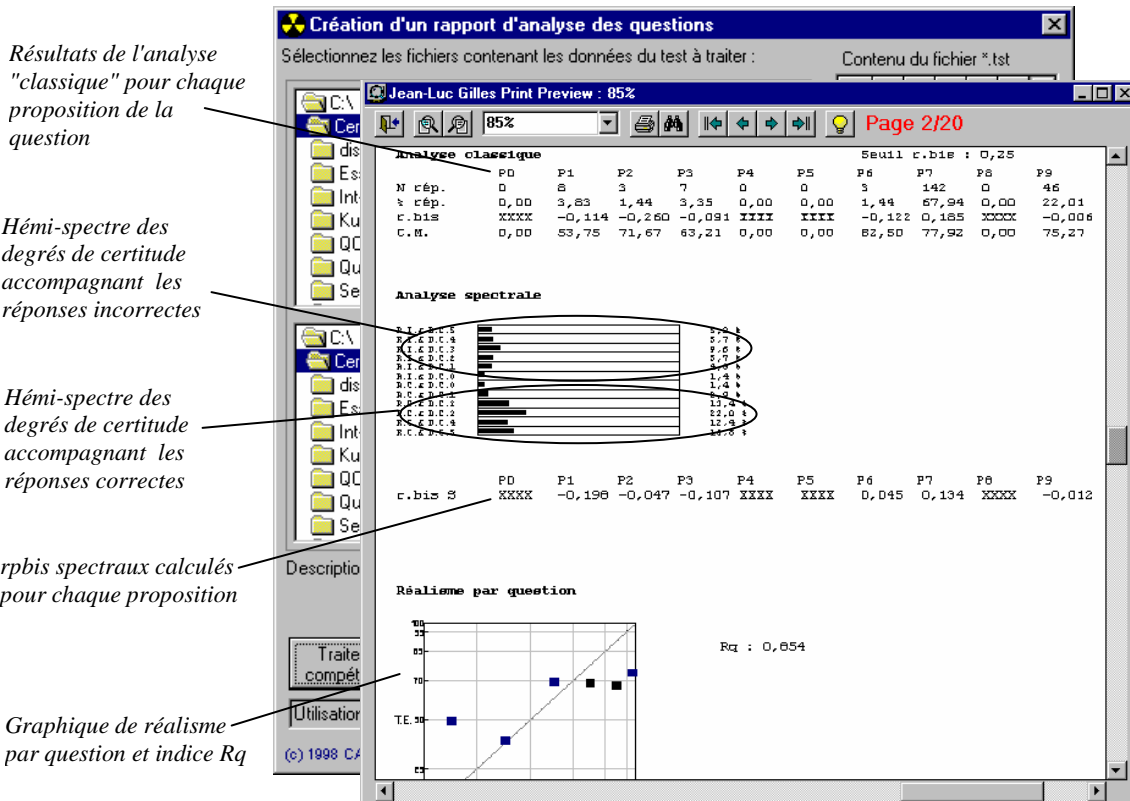
<sup>27</sup> CERT a été conçu en 1991 par le Service de Technologie de l'Éducation (STE) sous les auspices de la Commission des Communautés européennes dans le cadre du programme EUROTECNET (Boxus & al., 1991). Le logiciel a été programmé dans le langage Clipper par MM. Philippe de Brogniez et Henri Wuidar.

Il est par ailleurs aussi possible d'utiliser les informations fournies par les pourcentages de certitude qui ont accompagné les réponses des étudiants pour mesurer la cohérence spectrale des questions (Gilles, 1998a ; Gilles, 2002). Jusqu'ici, c'est un autre logiciel de traitement intitulé « SCANTEST 1.0 » qui a alors été utilisé.

## Le logiciel SCANTEST (version 1.0)

L'utilisation des informations liées à l'emploi des pourcentages de certitude en vue d'évaluer la qualité des QCM et QROC nous a amené à programmer un nouveau type de logiciel de traitement qui permet de calculer des rpbis Spectraux Contrastés (*rpbis SC*) et des rpbis Spectraux Contrastés après Turbo analyse (*rpbis SCT*) et donc d'effectuer une analyse spectrale de la qualité des épreuves standardisées (Gilles, 2002).

Voici un aperçu du premier logiciel que nous avons créé pour accélérer les traitements, il s'agit de la première version (1.0) de SCANTEST<sup>28</sup> (Gilles, 1998a).



Exemple de rapport de traitement pour une question - logiciel SCANTEST 1.0

Dans sa version 1.0, le programme SCANTEST utilise les fichiers des réponses des étudiants et de paramétrage des tests habituellement employés dans le cadre des traitements effectués à l'aide du logiciel CERT, ce qui permet de l'intégrer dans la chaîne de lecture optique de marque du SMART. Les sorties peuvent être visualisées à l'écran ou imprimées. Le rapport final comprend pour chaque question :

<sup>28</sup> Le logiciel SCANTEST 1.0 a été conçu et programmé par Jean-Luc Gilles à l'aide du langage MS Visual Basic 5.0.

- une analyse "classique" des propositions (ces résultats étaient déjà disponibles dans CERT): nombre d'étudiants (N rép.) qui ont choisi la proposition, pourcentage (% rép.), corrélation bisériale de point (*rpbis classique*) et certitude moyenne (C.M.) ;
- une analyse spectrale avec le graphique du profil spectral qui reprend pour chaque degré de certitude le pourcentage des étudiants qui ont choisi une proposition incorrecte (hémi-spectre des réponses incorrectes accompagnées des degrés de certitude 5, 4, 3, 2, 1, 0) et le pourcentage des étudiants qui ont répondu correctement (hémi-spectre des réponses correctes accompagnées des degrés de certitudes 0, 1, 2, 3, 4, 5) ;
- les *rpbis spectraux* (*rpbis S*) calculés en corrélant les degrés de certitude qui ont accompagné les réponses avec les choix/rejets des différentes propositions de la QCM;
- un graphique de Réalisation des prédictions par question (*Rq*) reprenant pour chaque degré de certitude (en abscisse) les taux d'exactitude correspondant (en ordonnée). Le graphique de chaque question est accompagné de la valeur de son indice *Rq*.

Des synthèses sont également proposées :

- un récapitulatif des certitudes moyennes pour les réponses correctes et incorrectes accompagné d'un graphique de cohérence métacognitive ;
- un récapitulatif des *rpbis classiques* pour les réponses correctes et incorrectes avec un graphique de qualité des *rpbis classiques* reprenant ces données pour l'ensemble des questions ;
- un récapitulatif des *rpbis spectraux* pour les réponses correctes et incorrectes accompagné d'un graphique de qualité des *rpbis S*.

Cependant dans cette première version du logiciel *SCANTEST* les *rpbis spectraux* n'étaient pas contrastés et le principe de la turbo analyse n'était pas encore implémenté. Ces fonctionnalités de traitement contrasté et de turbo analyse ont été introduites dans une nouvelle version (2.0) du logiciel *SCANTEST*<sup>29</sup> (Gilles, 2002) programmée en vue d'étudier et de mettre au point ces nouveaux instruments d'analyse spectrale de la qualité des épreuves. La nouvelle version de *SCANTEST* a été programmée pour permettre l'analyse spectrale des résultats des milliers d'étudiants (entre 1.410 et 3.846 sujets selon les tests) ayant participé aux 10 épreuves du projet de *MONITORING HISTORIQUE* de cohortes de *CANDIDATURES UNIVERSITAIRES* (*MOHICAN*).

### g) Feedbacks aux évalués après l'épreuve

Cette étape consiste en la communication aux étudiants, d'une part, des résultats des performances individuelles et, d'autre part, des résultats des contrôles de qualité relatifs aux questions (les performances de l'épreuve, c'est-à-dire des listings provenant du programme CERT (voir plus haut) que l'évaluateur annoté en expliquant les éventuelles mesures correctives qui ont été prises en vue d'améliorer la qualité de l'épreuve. Ces listings annotés peuvent ensuite être diffusés auprès des évalués, par exemple par voie d'affichage.

Nous verrons plus loin que lorsqu'on demande aux étudiants du premier cycle de la FAPSE-ULg leur avis sur la qualité des examens qu'ils ont subi, leurs reproches les plus marqués se situent au niveau du manque de rétroactions en ce qui concerne leurs performances aux examens. Ce problème a aussi été mis en avant par les étudiants de l'Université de Montréal dans le cadre

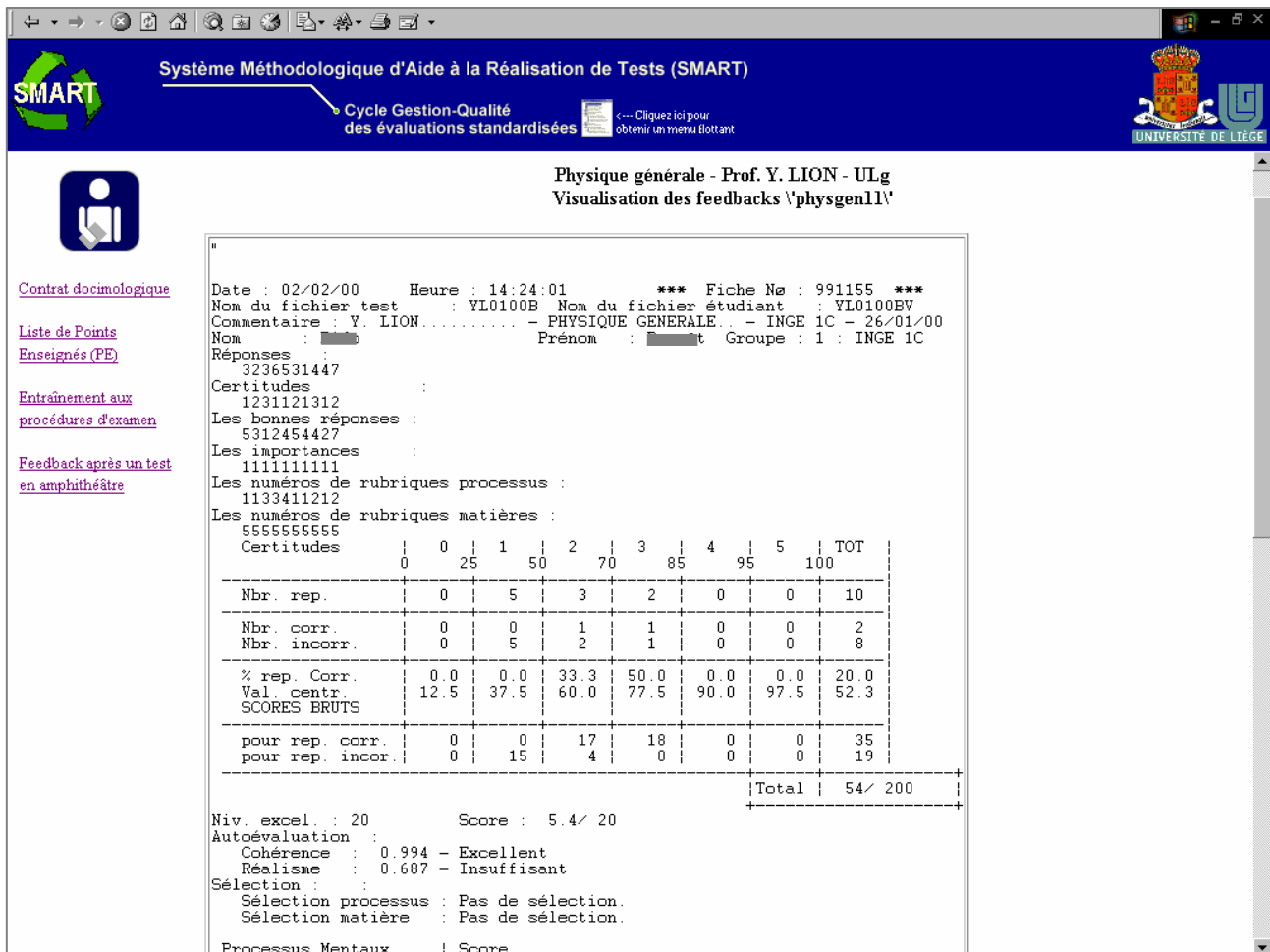
<sup>29</sup> Le logiciel *SCANTEST 2.0* a été conçu et programmé par Jean-Luc Gilles à l'aide de macro procédures en langage *Microsoft Visual Basic pour EXCEL*.

d'une enquête sur les pratiques d'évaluation des apprentissages menée par le Groupe de Recherche Interdisciplinaire en Pédagogie Universitaire (GRIPU) (Blais & al., 1997).

Cette équipe de chercheurs signale à propos des avis émis par les étudiants interrogés ( $n = 54$ ) : « Leurs demandes même si elles ne peuvent pas être généralisées, pointent des dimensions où une fragilité des pratiques se manifeste : le manque de rétroaction (absence de commentaires ou commentaires réduits ne permettant pas de comprendre les erreurs et d'apprendre de celles-ci)... » (Blais & al., p. 100).

Dans le but d'essayer d'améliorer la communication des résultats après les examens, nous avons mis en place depuis l'année académique 1998-1999 dans le cadre des épreuves standardisées réalisées à l'aide du dispositif de lecture optique de marques du SMART, une procédure d'accès aux feedbacks individualisés via l'Internet. Les étudiants peuvent obtenir leurs résultats après un examen standardisé après s'être connectés au site web du SMART (et après avoir fourni leur code d'accès).

Voici un aperçu de l'interface expérimentale proposée par le SMART et utilisée notamment avec les étudiants ingénieurs ayant participé aux interrogations hebdomadaires du cours de physique lors de l'année académique 1999-2000 (Lion, 2000).



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

Cycle Gestion-Qualité des évaluations standardisées

Physique générale - Prof. Y. LION - ULg  
 Visualisation des feedbacks 'physgen11'

Contrat docimologique  
 Liste de Points Enseignés (PE)  
 Entraînement aux procédures d'examen  
 Feedback après un test en amphithéâtre

Date : 02/02/00 Heure : 14:24:01 \*\*\* Fiche N° : 991155 \*\*\*  
 Nom du fichier test : YL0100B Nom du fichier étudiant : YL0100BV  
 Commentaire : Y. LION - PHYSIQUE GENERALE - INGE 1C - 26/01/00  
 Nom : Prénom : Groupe : 1 : INGE 1C

Réponses : 3236531447  
 Certitudes : 1231121312  
 Les bonnes réponses : 5312454427  
 Les importances : 1111111111  
 Les numéros de rubriques processus : 1133411212  
 Les numéros de rubriques matières : 5555555555

Certitudes	0	1	2	3	4	5	TOT
	0	25	50	70	85	95	100
Nbr. rep.	0	5	3	2	0	0	10
Nbr. corr.	0	0	1	1	0	0	2
Nbr. incorr.	0	5	2	1	0	0	8
% rep. Corr.	0.0	0.0	33.3	50.0	0.0	0.0	20.0
Val. centr.	12.5	37.5	60.0	77.5	90.0	97.5	52.3
SCORES BRUTS							
pour rep. corr.	0	0	17	18	0	0	35
pour rep. incorr.	0	15	4	0	0	0	19
<b>Total</b>							<b>54/ 200</b>

Niv. excel. : 20 Score : 5.4/ 20  
 Autoévaluation :  
 Cohérence : 0.994 - Excellent  
 Réalisme : 0.687 - Insuffisant  
 Sélection :  
 Sélection processus : Pas de sélection.  
 Sélection matière : Pas de sélection.  
 Processus Mentaux | Score

Le succès rencontré par cette possibilité de consulter son feedback après une épreuve nous fait penser que ce service offert par le SMART aux étudiants répond à un réel besoin chez ces derniers. Les statistiques de fréquentation du site web du SMART lors de la période de mise en œuvre de la première expérience de diffusion des feedbacks via l'Internet dans le cadre des interrogations hebdomadaires du cours de Physique destiné aux étudiants de 1<sup>ère</sup> candidature ingénieurs (de novembre 1999 à février 2000) ont montré une augmentation sensible du nombre de connexions au serveur en parallèle avec une augmentation de la consultation des pages du site qui permettaient l'accès aux feedbacks individualisés.

### h) Macro régulation

La dernière étape « Macro-régulation » permet d'organiser l'amélioration de tout le processus pour un cycle ultérieur de construction d'examen. C'est dans le cadre de cette 8<sup>ème</sup> étape que sont recueillis les avis des évalués dans le contexte du SMART.

A l'Université de Liège, depuis 1997, une Commission interfacultaire d'EVALUATION des ENSEIGNEMENTS (EVALENS) mise en place par le Conseil Général des Etudes (CGE) organise à l'échelle de l'institution une évaluation systématique des cours, travaux pratiques et examens. De 1997 à 1999, au sein de chaque faculté, des sous-commissions ont été chargées de rédiger les items qui devaient figurer sur les questionnaires d'avis (les questionnaires ont donc été créés sur mesure pour chaque faculté durant cette période). Depuis le début des opérations, le SMART a été sollicité pour fournir une aide méthodologique et logistique à la Commission EVALENS. Des formulaires spéciaux permettant le dépouillement par lecture optique de marques des avis des étudiants ont ainsi été conçus en étroite collaboration avec les facultés.

A la Faculté de Psychologie et des Sciences de l'Education (FAPSE-ULg), trois types de questionnaires furent proposés par la commission facultaire d'évaluation des enseignements : un premier pour les cours, un second pour les travaux pratiques et un troisième pour les examens. Dans cette partie nous allons d'abord exposer la façon dont les avis des étudiants ont été récoltés et traités en vue d'améliorer la qualité des examens à la FAPSE-ULg ainsi que le mode de calcul des moyennes des avis. Ensuite nous présenterons les faits saillants liés aux avis récoltés depuis 1997.

#### *Procédure de recueil des avis des étudiants à propos des examens à la FAPSE-ULg*

A la fin du dernier examen de la 1<sup>ère</sup> session en juin 1998 chaque étudiant inscrit en 1<sup>ère</sup> ou en 2<sup>ème</sup> candidature reçut une série de 15 *formuloms* permettant aux étudiants de donner leur avis à propos des examens en répondant à 13 items. Pour des raisons budgétaires, en juin 1999 l'opération eu lieu uniquement en 1<sup>ère</sup> candidature. Les questionnaires furent complétés sur place (un *formulom* par examen évalué). Chaque *formulom* contenait 13 items qui étaient répartis dans 3 catégories : « 1. Mode d'évaluation », « 2. Attitude de l'examineur » et « 3. Feedback après examen ». Le choix des 13 items présentés dans les *formuloms* fut le résultat d'un compromis au niveau :

- du nombre maximum d'items auxquels il était possible de répondre en un temps raisonnable à la fin du dernier examen (13 questions \* 15 cours cela représente 195 items à cocher) ;
- du nombre minimum de questions à poser pour donner une vue d'ensemble suffisamment détaillée de la façon dont les étudiants percevaient les examens (tous les objectifs « qualité » que nous avons présentés ne sont malheureusement pas couverts par ces 13 items, ni d'ailleurs toutes les

8 étapes du cycle de réalisation d'une épreuve) ;

- de l'importance à accorder aux différentes questions possibles dans le contexte facultaire (dans le contexte FAPSE, le choix des 13 questions sélectionnées fut discuté entre les membres de la Commission facultaire EVALENS, puis soumis au Conseil des études et ensuite voté en Conseil de faculté).

Voici les 13 items utilisés en 1997-1998 et 1998-1999 :

**1. Mode d'évaluation**

- [1.1] Le mode d'évaluation (QCM, oral, écrit, travaux personnels) était adéquat
- [1.2] L'entraînement à la procédure d'évaluation avant l'examen était suffisant
- [1.3] Les questions d'examen étaient clairement formulées
- [1.4] Les questions d'examen étaient bien adaptées à la matière
- [1.5] Le mode d'évaluation choisi permet au professeur d'avoir une bonne représentation des compétences acquises par l'étudiant
- [1.6] L'évaluation est équitable et impartiale

**2. Attitude de l'examineur**

- [2.1] Les exigences de l'enseignant sont clairement présentées aux étudiants
- [2.2] Les exigences de l'enseignant sont présentées en temps utiles
- [2.3] L'enseignant met l'étudiant à l'aise à l'examen oral (si écrit, indiquez SO)
- [2.4] L'horaire fixé pour l'examen est respecté

**3. Feedback après examen**

- [3.1] Les réponses correctes sont communiquées aux étudiants après l'examen
- [3.2] L'analyse statistique de la qualité des questions (r.bis) est communiquée
- [3.3] Après l'examen, l'étudiant peut obtenir des explications sur la qualité de ses réponses auprès de l'enseignant.

Chaque item se présente donc sous la forme d'une affirmation relative à l'examen évalué. L'étudiant exprime son avis en se référant à l'échelle ci-dessous. Cette échelle est reprise sur le *formulom* où il coche son choix.

SO = sans objet	1 = pas du tout d'accord	2 = pas d'accord	3 = plutôt pas d'accord	4 = plutôt d'accord	5 = d'accord	6 = tout à fait d'accord	SA = sans avis
-----------------	--------------------------	------------------	-------------------------	---------------------	--------------	--------------------------	----------------

Après lecture optique et traitement des avis des étudiants deux types de feedbacks sont diffusés. Le premier qui est public et anonyme est destiné à être affiché aux valves de la faculté. Voici un exemplaire de ces données publiques relatives à l'évaluation des examens par les étudiants de 1<sup>ère</sup> candidature en 1997-1998.

**Faculté de Psychologie et Sciences de l'Education - 1<sup>ère</sup> candidature (1997-1998)**

**Tableau des moyennes (anonymes) pour l'ensemble des examens de la section évaluée**

Les chiffres qui figurent dans les cases indiquent le nombre d'examens de cette section qui récoltent la moyenne correspondant à la case.

		Moyennes des avis					
		1	2	3	4	5	6
		1,5	2,5	3,5	4,5	5,5	
<b>1.</b>	<b>Mode d'évaluation</b>						
1.1	Le mode d'évaluation (QCM, oral, écrit, travaux personnels) était adéquat			2	2	9	2
1.2	L'entraînement à la procédure d'évaluation avant l'examen était suffisant		1	3	5	1	5
1.3	Les questions d'examen étaient clairement formulées			2	5	6	2
1.4	Les questions d'examen étaient bien adaptées à la matière			2	3	8	2
1.5	Le mode d'évaluation choisi permet au professeur d'avoir une bonne représentation des compétences acquises par l'étudiant			3	7	4	1
1.6	L'évaluation est équitable et impartiale		1	3	4	6	1
<b>2.</b>	<b>Attitude de l'examineur</b>						
2.1	Les exigences de l'enseignant sont clairement présentées aux étudiants			1	3	4	6
2.2	Les exigences de l'enseignant sont présentées en temps utiles			1	3	3	6
2.3	L'enseignant met l'étudiant à l'aise à l'examen oral (si écrit, indiquez SO)			4	7	4	
2.4	L'horaire fixé pour l'examen est respecté				2	9	4
<b>3.</b>	<b>Feed-back après examen</b>						
3.1	Les réponses correctes sont communiquées aux étudiants après l'examen	1	8	2	3	1	
3.2	L'analyse statistique de la qualité des questions (r.bis) est communiquée	1	10	2	1	1	
3.2	Après l'examen, l'étudiant peut obtenir des explications sur la qualité de ses réponses auprès de l'enseignant.		4	5	4	2	

A droite des items on remarque une série de 10 cases. Dans le cas du premier item « 1.1 *Le mode d'évaluation (QCM, oral, écrit, travaux personnels) était adéquat* » la cinquième case comprise entre 3 et 3,5 sur l'échelle proposée contient un « 2 ». Cela signifie que 2 enseignements de 1<sup>ère</sup> candidature récoltent une moyenne située entre 3 et 3,5 à cet item. Etant donné qu'il s'agit ici de données publiques (anonymes) on ne peut savoir de quels enseignements il s'agit.

Dans la procédure EVALENS, cette synthèse des avis relatifs aux examens (ainsi que celles des cours et travaux pratiques) est non seulement affichée aux valves de la faculté, mais aussi présentée lors d'un Conseil de faculté où tous les membres du corps académique ainsi que les représentants du corps scientifique et les délégués des étudiants sont présents et peuvent intervenir pour obtenir des clarifications ou émettre des suggestions.

Le second type de feedback est personnel et destiné à chaque enseignant. C'est le même que celui qui figure ci-dessus à la différence que la position de l'enseignant est signalée par une case grisée. Chaque enseignant reçoit son feedback personnel sous pli fermé et une copie est envoyée au Doyen de la faculté. Ce dernier organise une discussion à propos des avis des étudiants avec chaque professeur.

En plus de sa synthèse personnelle, le professeur reçoit les formulaires sur lesquels les étudiants ont noté leurs commentaires dans la case « *Suggestion(s) à propos de l'évaluation du cours* ».

### Mode de calcul des moyennes pour l'ensemble des examens d'une section

La moyenne des avis récoltée par chaque examen à chaque item du formulom est calculée de la façon suivante. Voici un exemple (fictif) de répartition des avis dans le cadre d'un item pour lequel 235 étudiants ont donné leur avis :

SO = sans objet	1 = pas du tout d'accord	2 = pas d'accord	3 = plutôt pas d'accord	4 = plutôt d'accord	5 = d'accord	6 = tout à fait d'accord	SA = sans avis
	1	2	3	4	5	6	
n = 5	n = 10	n = 30	n = 35	n = 70	n = 55	n = 15	n = 15
Seuls les avis de « 1 » à « 6 » sont pris en compte dans le calcul de la moyenne (n = 215)							

Le tableau montre que 5 étudiants ont coché « SO = sans objet », 10 ont coché « 1 = pas du tout d'accord », 30 ont coché « 2 = pas d'accord », etc. Au total, pour cet enseignement, 215 avis ont été retenus après décompte des SO et SA. Les avis « 1 = pas du tout d'accord » à « 6 = tout à fait d'accord » sont considérés comme des points accordés à l'enseignement par les étudiants, ce qui donne :

$$(10 * 1) + (30 * 2) + (35 * 3) + (70 * 4) + (55 * 5) + (6 * 15) = 820$$

Dès lors, la moyenne récoltée par l'examen à cet item vaut  $820/215 = 3,8$  (le minimum pourrait être 1 et le maximum 6). Remarquons qu'étant donné l'échelle d'avis proposée, la note charnière se situe à 3,5 c'est-à-dire entre le négatif « 3 = plutôt pas d'accord » et le positif « 4 = plutôt d'accord ».

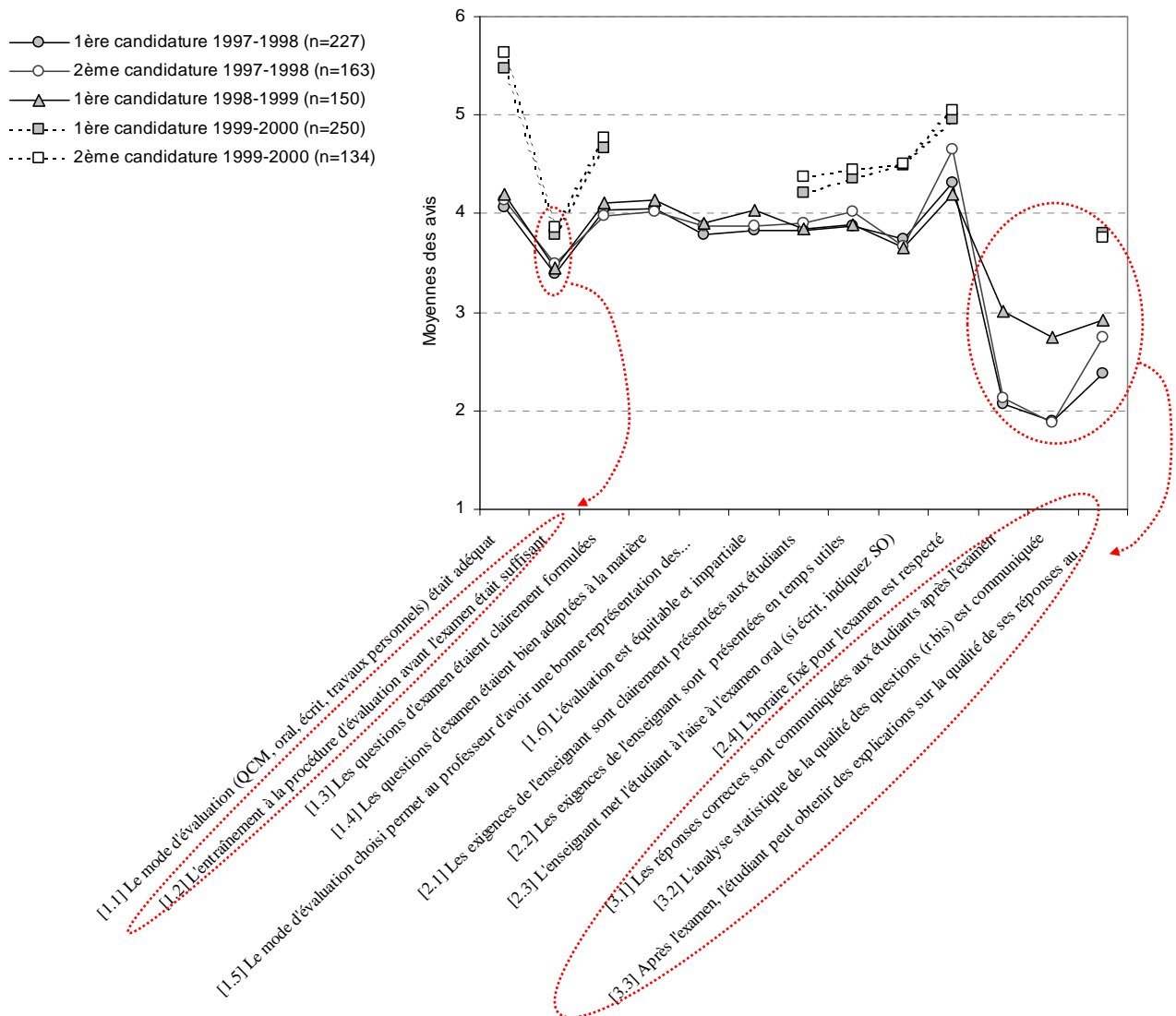
#### ***Faits saillants liés aux avis récoltés depuis 1997 à propos des examens à la FAPSE***

Le graphique ci-dessous reprend les moyennes des avis des étudiants calculées pour l'ensemble des examens organisés en 1<sup>ère</sup> candidature ou/et en 2<sup>ème</sup> candidature à la FAPSE-ULg de 1997 à 2000 (Gilles & al., 1998). Chacun des 13 items rappelés sous le graphique reçoit une moyenne qui peut aller de 1 à 6. Les points représentant les moyennes obtenues par une candidature ont été reliés afin d'aider le lecteur à visualiser les résultats section par section.

En 1999-2000, la procédure de récolte des avis fut modifiée ainsi que le questionnaire qui ne comportait plus tous les items proposés les deux années académiques précédentes (5 items furent supprimés et 11 nouveaux items furent ajoutés). Ceci explique l'absence de carrés (et de traits pointillés) pour les deux candidatures relatives à 1999-2000 en ce qui concerne les items 1.4 à 1.6, 3.1 et 3.2. En 1998-1999, la 2<sup>ème</sup> candidature ne fut pas évaluée (pour des raisons budgétaires).

Les « n » qui figurent dans la légende représentent les nombres moyens d'étudiants qui ont donné leurs avis sur les examens dans les différentes sections envisagées. Nous remarquons que pour la 1<sup>ère</sup> et la 2<sup>ème</sup> candidature de l'année 1997-1998 (ronds grisés et blancs reliés par des traits

Comparaison des moyennes des avis des étudiants à propos des examens dans le 1er cycle de la FAPSE-ULg (de 1997 à 2000)



continus) les courbes des moyennes des avis sont très proches. La courbe des avis récoltés en 1<sup>ère</sup> candidature en 1998-1999 (triangles grisés et trait continu) est aussi fort proche de ces deux dernières.

Globalement, nous remarquons que les items qui récoltent les moyennes d'avis les moins favorables sont liés d'une part à l'entraînement à la procédure d'évaluation avant l'examen et d'autre part aux feedback après l'examen. Ces résultats montrent que les étudiants ne se sentent pas suffisamment préparés aux modalités d'évaluation et de façon encore plus marquée que les feedbacks qu'ils peuvent obtenir après l'examen sont insuffisants. On constate cependant une nette amélioration en 1998-1999 par rapport à l'année précédente en ce qui concerne les 3 derniers items relatifs à la catégorie « 3. Feedback après l'examen ».

Rappelons que ces résultats sont présentés et discutés en Conseil de faculté et que des entrevues individuelles sont organisées entre chaque professeur et le Doyen de la faculté lors desquelles des actions pour améliorer la situation sont envisagées. Les enseignants peuvent alors s'expliquer à propos des moyennes que récoltent leurs examens et c'est aussi pour eux l'occasion d'exprimer des besoins qui permettraient d'augmenter la qualité de leurs épreuves.

On remarque en 1999-2000 une amélioration aux huit items repris des questionnaires d'avis présentés les années précédentes. Lorsqu'on compare les moyennes obtenues aux huit items, on remarque que les deux items relatifs à l'entraînement et au feedback après examen restent ceux qui obtiennent les avis les moins favorables.

### ***Faits saillants pour l'année académique 1999-2000***

Précédemment nous avons signalé que le questionnaire utilisé en 1999-2000 était différent. Voici les 19 items relatifs aux examens qui furent proposés aux étudiants.

#### **1. Préalables à l'examen**

- [1.1] L'entraînement à la procédure d'évaluation avant l'examen était suffisant
- [1.2] Les exigences de l'enseignant sont clairement présentées aux étudiants
- [1.3] Les exigences de l'enseignant sont présentées aux étudiants en temps utile

#### **2. Mode d'évaluation :**

- [2.1] QCM : ce mode d'évaluation était adéquat
- [2.2] Certitudes : ce mode d'évaluation était adéquat
- [2.3] Examen oral : ce mode d'évaluation était adéquat
- [2.4] Examen écrit rédigé : ce mode d'évaluation était adéquat
- [2.5] Présentation écrite d'un travail (rapport, fiche de lecture, ...) : ce mode d'évaluation était adéquat

#### **3. Organisation de l'examen**

- [3.1] L'horaire fixé pour l'examen est respecté
- [3.2] La durée de l'examen est suffisante
- [3.3] La surveillance anti-fraude est suffisamment dissuasive

#### **4. Qualité du questionnement**

- [4.1] Les questions de l'examen sont clairement formulées
- [4.2] La procédure d'examen recouvre bien toute la matière

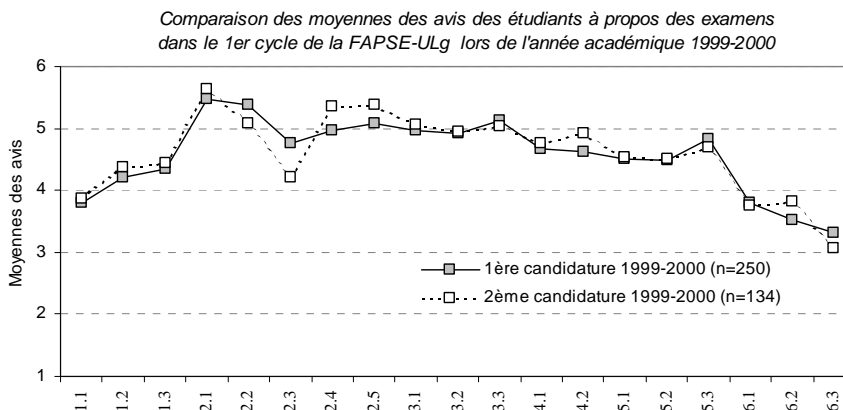
#### **5. L'examineur(trice)**

- [5.1] Adopte une attitude correcte à l'égard de l'étudiant, lors de l'examen oral
- [5.2] Met l'étudiant à l'aise lors de l'examen
- [5.3] Est impartial

#### **6. Après l'examen ...**

- [6.1] L'étudiant peut obtenir auprès de l'enseignant des explications sur la qualité de ses réponses
- [6.2] Les résultats (points ou grades) sont affichés
- [6.3] Pour les examens écrits, les réponses correctes sont communiquées dans les délais

Les points des courbes ci-contre montrent les moyennes des avis des étudiants récoltées par les 19 items en 1999-2000. Nous observons deux courbes très proches en 1<sup>ère</sup> et 2<sup>ème</sup> candidature (comme en 1997-1998).



On remarque à nouveau que l'item lié à l'entraînement (1.1) et ceux qui sont liés aux rétroactions après l'examen (6.1, 6.2 et 6.3) obtiennent les moyennes les plus faibles. Nous constatons aussi que l'item qui obtient la moyenne la plus élevée en 1<sup>ère</sup> et en 2<sup>ème</sup> candidature est celui qui concerne l'adéquation du mode de questionnement par QCM. Les certitudes sont aussi considérées comme un mode d'évaluation adéquat. En 1<sup>ère</sup> candidature les examens écrits rédigés et les travaux écrits rédigés sont considérés moins adéquats qu'en 2<sup>ème</sup> candidature. La moyenne obtenue par l'examen oral est quant à elle la moins élevée en ce qui concerne l'adéquation des modes d'évaluation, avec des différences entre la 1<sup>ère</sup> et la 2<sup>ème</sup> candidature dans la mesure où les étudiants de 2<sup>ème</sup> trouvent l'examen oral moins adéquat que les étudiants de 1<sup>ère</sup>.

### III. PRESENTATION DE L'ETAT DE LA RECHERCHE E-C&QCST A LA 11EME CONFERENCE BIENNALE DE L'EUROPEAN ASSOCIATION FOR RESEARCH IN LEARNING AND INSTRUCTION (EARLI)

#### 1. electronic Construction and Quality Control in Standardised Testing: the e-C&QCST project

*Communication présentée par Jean-Luc GILLES lors du congrès international « EARLI 2005 – 11th Biennial Conference of the European Association for Research in Learning and Instruction (EARLI) »*

*Auteurs : Jean-Luc GILLES<sup>1</sup>, Sylvie-Anne Piette<sup>2,3</sup>, Pascal Detroz<sup>3</sup>, Salvatore TINNIRELLO<sup>4</sup>, Marc PIRSON<sup>5</sup>, Mbassa DABO<sup>6</sup> et Hung LÊ<sup>7</sup>*

[PowerPoint en annexe I]



1. Chargé de recherches, directeur du Système Méthodologique d'Aide à la Réalisation de Tests (SMART) de l'Université de Liège - Belgique
2. Coordinateur des recherches en gestion des compétences – Unité d'Enseignement et de Recherches en Ressources Humaines – HEC Liège - Belgique
3. Coordinateur des évaluations des enseignements au Système Méthodologique d'Aide à la Réalisation de Tests (SMART) de l'Université de Liège - Belgique
4. Chercheur en gestion des compétences – Unité d'Enseignement et de Recherches en Ressources Humaines – HEC Liège - Belgique
5. Coordinateur de l'équipe informatique du projet e-C&QCST au Système Méthodologique d'Aide à la Réalisation de Tests (SMART) de l'Université de Liège - Belgique
6. Informaticien du projet e-C&QCST au Système Méthodologique d'Aide à la Réalisation de Tests (SMART) de l'Université de Liège - Belgique
7. Informaticien du projet e-C&QCST au Système Méthodologique d'Aide à la Réalisation de Tests (SMART) de l'Université de Liège - Belgique

## **eLECTRONIC CONSTRUCTION AND QUALITY CONTROL IN STANDARDIZED TESTING: THE e-C&QCST PROJECT**

Jean-Luc Gilles, Sylvie-Anne Piette, Pascal Detroz, Salvatore Tinnirello, Marc Pirson, Mbassa Dabo and Hung Lê

**University of Liège, Belgium**

### **a) Introduction**

The docimology research showed that the standardized testing is an efficient evaluation form of the learners' competencies provided to meet qualities of validity, reliability, sensitivity, diagnosticity, equity, practicability, communicability and authenticity. When standardized tests meet these qualities, they are appreciated by trainers, learners and by heads of training institutions. Currently, the construction and the quality management of standardized qualifying evaluations is a worry for the higher/university education in many European countries which have to face a still growing number of students. The training professionals and the human resources managers of companies are often constrained to evaluate staff competencies acquired outside the schools. Moreover, these concerns, regarding the quality of the tests, are also in relation with a worldwide movement of the introduction of quality management in education and training activities. Our research aims at facilitating the creation of standardized tests to make them more reliable, more efficient and quicker to set up, in order to satisfy at best the actual requirements of the education/training/human resources world. Our objective is to create an electronic platform of construction and quality control in standardized testing entitled "e-C&QCST (electronic Construction & Quality Control in Standardized Testing) build in an "Open Source" policy for a web based utilisation.

***Keywords:*** *Assessment; Standardized testing; Computer application*

The standardized testing is an evaluation form of the learners' competencies which - when it meets the qualities of validity, reliability, sensitivity, diagnosticity, equity, practicability, communicability and authenticity (Kirkland, 1971; Kulik & Kulik, 1987; Leclercq & Gilles, 1993; Gilles & Leclercq, 1995; Castaigne, Gilles & Hansen, 2001; Gilles, 2002; Phelps, 2005) – is appreciated by the trainers, learners and the heads of training institutions.

The preoccupations concerning the quality of standardized evaluations of learners' competencies fit into a movement which has been reaching the Belgian and European institutions in the early 90' (Kaufman & al., 1995; Segers & Dochy, 1996). This movement itself is part of a worldwide trend of introducing the quality management in the education and training activities spheres (Ramsden, 1991; Nightingale & O'Neil, 1994; Zink & Schmidt, 1995; Cascallar, Dochy & Segers, 2003).

The team of the unit entitled "Système Méthodologique d'Aide à la Réalisation de Test (SMART)" of the University of Liège helps academic teachers and trainers of private companies to construct assessments tools for measuring students' or trainees' competencies. The SMART proposes a quality approach for the development of standardized tests for ducational or training purposes. This quality approach is based on four fundamental principles which to be taken into account in the creation of evaluations (Gilles, 2002).

This approach is inspired by the recommendations of the international standard ISO 9004 2 (Quality management and quality system elements - Guidelines for the services). In the SMART context of docimologic engineering, the quality approach suggested allows to focus the actions on four types of concerns to be undertaken in the creation of learner's competencies tests. At the heart of our approach, the central principle is related to the needs of the developers, of the teachers/trainers, of the students/trainees and of the persons in charge of the evaluation tools which have to be conceived and developed. It is the central pole of our approach and also the point of convergence of three other poles: the pole of the models used as references for the development of the assessment tools, comprising the model entitled CCQCST "Cycle of Construction and Quality Control for Standardized Testing" (briefly exposed later); the pole of human and material resources to mobilize and the pole of the quality standards which have to guide us in our actions.

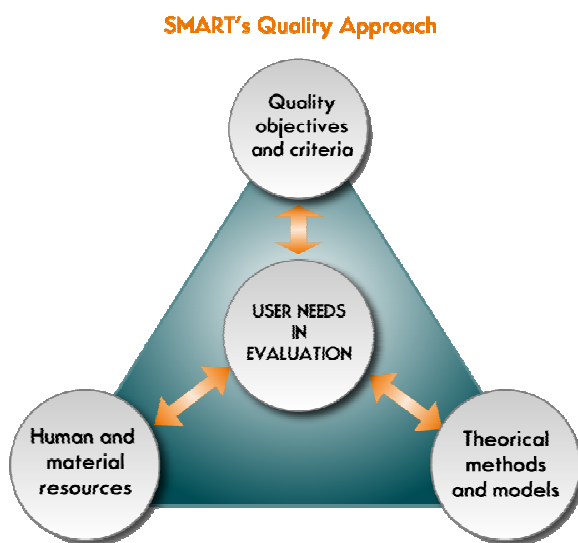


Figure 1. SMART's quality approach in standardized testing construction

In this context, our research aims at setting up methods, procedures and tools for the quality management of standardized tests. The spreading and the worldwide utilisation of our research results are possible today thanks to interfaces integrated in the software we propose. They are indeed created to work in a web based environment. This is the idea at the base of our research and development project entitled "electronic Construction and Quality Control in Standardized Testing (e-C&QCST)". The e-C&QCST platform is developed in an "Open Source" policy.

Our contacts with private institutions within the context of our research confirm our hypothesis of a large market which opens for the e-C&QCST platform and for the service companies which would ensure a coaching and a follow up of the users.

Furthermore the idea of a quality label for standardized testing (Gilles, 1998) will be proposed in the perspective of a qualitative docimology in adequacy with the trainers' and trainees' needs. This "labellisation" perspective seems to meet some expectations in the training world. In this economic valorisation perspective of the e-C&QCST research, we propose to structure the supply of services in the following way:

[1] Audit of the standardized testing systems of a client company/organization to measure the adequacy of the human and material resources, the needs of the client, his objectives and the expected results;

[2] Training and coaching of the users of the e-C&QCST platform within their company/organization context in order to facilitate the appropriation and integration of the platform in the training and evaluation system;

[3] Docimological services for the improvement of the standardized tests of a client company/organization. These kinds of services should be facilitated by the use of the e-C&QCST platform which permits to a person in charge of the development of a standardized test to give partial access to subcontractors for precise tasks;

[4] Tests creation with our e-C&QCST platform for third party companies/organizations.

### **b) Method**

#### ***The modules of the e-C&QCST platform***

The idea is to formalize, adapt or create docimological procedures in order to control and quantify the quality of standardized tests of learners'/trainees' competencies. For this purpose, we will use the Cycle of Construction and Quality Control in Standardized Tests (CCQCST) model that we have proposed (Gilles et Leclercq, 1995; Gilles, 2002). The 8 realisation steps of this model will be programmed in the web based e-C&QCST platform. This platform is being created with the free programming language PHP and the free database manager MySQL. We also use the free server Apache.

Since the middle of the nineties, this eight steps have been constantly improved by the SMART team.

The 8 steps of the CCQCST model related to the 8 modules of the e-C&QCST platform are:

- [1] ANALYSIS: The course/training or competencies description must be analysed in order to list out the points to evaluate (PE), their priorities (P) and the categories of performances (CP) of the learners/trainees. The end product consists of a test specification table where the points to evaluate and their priorities are crossed with the performances categories (PE x CP).
- [2] DESIGN: In the second step of construction the developer must design the multiple choice questions modalities and make decisions about the importance of each modality. The end products are of two types. On one hand, there must be an association of categories of performances of the previous step with the modalities of the questions (CP x MQ) defined at this step (i.e. a MQ could be a type of MCQs with confidence degrees). On the other hand, the binomials (PE x CP) and (CP x MQ) must be combined to obtain trinomial ones (PE x CP x MQ). Finally, from the list of the trinomials objects the ones which should be linked to questions during the test must be selected.
- [3] QUESTIONS: In the third step, the test developer selects or writes the questions which are linked to the trinomials objects selected for the test at the former step of the construction. The questions are created or selected in a secured item bank database.
- [4] TRAINING: At the fourth step, the person in charge of the development of a test can choose to propose "blank tests" (paper-pencil or web based) in order to train the evaluated persons on the modalities of the questions previously defined.

- [5] **TESTING:** In the fifth step, the questions created or selected at the third step are put in the standardized test which is proposed to the people who perform the test (paper-pencil or web based). In certain cases, it is possible to perform online exams - in a secured environment - or to link the questionnaires to automated responses acquisition devices such as optical marks reading.
- [6] **CORRECTION:** At the sixth step, the standardized test developer carries out the correction of the tests. Treatments are designed to check the quality of the questions before the final version of the results of the evaluated persons is validated in one form or another.
- [7] **FEEDBACK:** Personalized and tailored feedbacks with for example online access are proposed on the basis of the final version of the results proposed at the previous step.
- [8] **REGULATION:** This is the eighth and last step of the construction of the standardized test. Here, the test developer collects and analyses the opinions of the all the people who intervened in the test construction process. The aim is to improve the next standardized test which will be created. This explains and justifies the C of the term “Cycle” in the acronym “CCQCST” of our model.

For each construction step defined above, quality control procedures are also considered within the framework of the research and development of the e-C&QCST platform.

### Quality criteria for standardized testing construction

The docimological procedures used in the 8 modules of the e-C&QCST platform are built in reference to a series of quality criteria that are used by the SMART team in standardized testing construction (Gilles, 1998, 2002; Gilles et Lovinfosse 2004):

- **Validity:** candidates' scores must reflect what the trainer/teacher wants to measure. The idea is to ensure solid inferences with the categories of performances (CP, construct or theoretical validity) and to cover the important aspects of the learning/training, what we called the points to be evaluated (PE, content validity);
- **Fidelity:** an evaluated person must essentially obtain the same score if his/her copy is corrected by different persons in charge of the correction (inter-judge corroborating) and/or at another moment by the same person or the same team (intra-judge consistency);
- **Sensibility:** the candidate's competencies measures must be precise and reflect subtle phenomena;
- **Diagnosticity:** tests results must ensure a precise diagnostic of the learning difficulties (and ideally of their causes) and of the competencies which are mastered and those which are not;
- **Practicability:** the feasibility in terms of time, human and material resources must be ensured;
- **Equity:** each candidate must be treated in a fair way, as far as possible in the same way;
- **Communicability:** non confidential information concerning the process progress must be communicated and understood by the people who perform the test;
- **Authenticity:** the questions must be in relation with situations which make sense for the people who perform the test, relevant in comparison with the real world and able to enhance transfers in contexts and practices that can be met in life.

### *Roles' distribution in the research and development team*

The SMART team, who is the promoter of this e-C&QCST project, joined for the needs of this research the team of Management of Competencies of the High School of Management at the University of Liege. The General Direction of Technologies, Research and Energy of the Walloon area of Belgium supported financially this R&D project since the end 2003 for a period of three years.

The roles in the research group are distributed in the following way. The SMART team is divided into two subgroups. The first subgroup - two part time scientists: one specialized in docimology (he coordinates the research) and the other in ergonomics - conceives and formalizes the docimological procedures used within the framework of the CCQCST model and writes the project specifications of the e-C&QCST platform. The second subgroup (three data processing specialists, two full-time and one part time) programs the project specifications of the eight modules, starting with the first and finishing by the 8th. The Management of Competencies team (one half-time economist and a 10 percent full-time equivalent scientist specialized in human resources management) organizes the contacts with the companies/organizations which are partners in this project. In particular this Management of Competencies team submits the scenarios of the docimological procedures to the companies/organizations in order to validate the procedures and accompanies the partners during the experimentations of the eight modules. These tests of the modules in the world of the companies as well as other tests realized in the academic world bring readjustments of the project specifications. They are discussed within the research group.

### *Research and development hypotheses*

This project covers various conceptual frameworks to which we attach research and development hypotheses. We identified eight conceptual frameworks and propose these principal hypotheses:

- (A) Docimological framework: the quality of standardized tests could be optimized by combining different tests construction procedures at the eight steps of the CCQCST model which will make it possible to guarantee a high quality level at the eight criteria we have defined (validity, fidelity, sensitivity, diagnosticity, feasibility, equity, communicability and authenticity). The requirements in standardized testing of the companies' world are sufficiently close to those of the academic world so that it is possible to create an e-C&QCST platform which can be appropriate in both cases.
- (B) Technological framework: an electronic platform covering all 8 steps of the CCQCST model helps to improve the quality when assessment developers create standardized tests.
- (C) Framework of ergonomic aspects: it is possible to develop an electronic platform which makes it possible to implement the Cycle of Construction and Quality Control in Standardized Tests (CCQCST) model in an intuitive way.
- (D) Framework of quality management: quality controls are possible in link with our quality criteria at the different steps of the CCQCST model. The use of these quality controls increases the quality of the standardized tests. These quality controls could be used in a quality label policy.
- (E) Economical valorisation framework: there is a market for subcontracting services in the field of an economical valorisation of the e-C&QCST platform. There is also a market for the development of a quality label related to the standardized testing.

- (F) Users framework: it is possible to develop a community of users of the e-C&QCST platform who could develop interactions by sharing practices, by exchanging tests... Those interactions could improve the competencies in quality standardized tests construction of tests developers. The use of the e-C&QCST platform by companies/organizations changes the vision of the standardized testing and contributed to the emergence of the complementarities and the responsibilities of the roles of the persons who are involved in test creation.
- (G) Linguistic framework: a broad diffusion of the e-C&QCST platform implies a translation of the interfaces in most languages of the European Union.
- (H) Experimentations framework: a collaborative design with a sample of companies/organizations, future potential users, implied in each construction step (the eight modules of the e-C&QCST platform), makes it possible to check the adequacy of the procedures suggested with the users' needs. Collaborative design makes it also possible to consider different uses from those initially envisaged and to progressively test the software ergonomics while the product is constructed.

In this article, we will discuss first results concerning the hypotheses from the frameworks A, C, E and F. At this stage, at middle of the research, it is still too early to learn all the useful lessons concerning the others frameworks hypotheses.

### c) First results

#### *Companies and organizations collaborating in this research*

As we saw above, the SMART created the CCQCST model and made use of it since middle of the nineties within the framework of the help provided to higher education teachers. Former studies showed the effectiveness of this model in the academic context (Gilles & Leclercq, 1995; Castaigne & Gilles, 2000; Gilles, 2002; Gilles & Lovinfosse, 2004). But this model was used by few trainers of private companies and organizations. This is the reason while early in this project we proposed collaborations to private companies or organizations.

We contacted the EPSILON association to help us in this task. This association proposes to all the private companies or organizations of Belgium to share their practices in relation with the development of the competencies of their collaborators. With the help of the EPSILON staff, we defined sample characteristics of the companies/organizations to be approached in order to propose to them to join the e-C&QCST project partnership. The starting idea was to meet people in charge of standardized tests in widely different companies and organizations. We knew that very small companies or organizations had few chances of being potential users of the e-C&QCST platform (excepting those whose trade is in training or in testing).

The characteristics of the companies and organizations selected to create the partnership in this R&D project were related to:

- The type of company (public, private, mixed);
- The size of the company or the organization and its network (membership or not of a group, great number of subcontractors...);
- The existence of an internal training service;
- The stakes of the evaluations (attesting, recruitment, mobility, career management...).

At the end of the selection process, we started collaborations with ten companies and organizations. These are:

- The FORTIS Bank;
- The AXA Bank and insurance group;
- The Training Institute of the Belgian's Federal Administrations (IFA) ;
- The Pizza Belgium company;
- The communal association of economic development of the Belgian province of Luxembourg (IDELUX);
- The FOREM (Belgian community and regional office of the vocational training and employment);
- The PHONE LANGUAGE company;
- The TNT Express Worldwide Euro HUB;
- The UNIQUE interim company;
- The "Consortium de validation des competences", a consortium of the Belgian french speaking community, the Walloon area and the Brussels area (Its mission is to create centers of validation of competencies in connection with the continuous vocational training)

### *Interest in the CCQCST model and its docimological procedures*

One of the first activities proposed to our partners consisted in presenting in details the 8 steps of the cycle of construction and quality control for standardized testing. Many partners expressed a great interest for the procedures we proposed. Four partners (AXA, FOREM, FORTIS Bank and IFA) requested immediate implementations of our docimologic procedures.

These requests for immediate docimological subcontracting services were carried out by members of the SMART team in parallel with the R&D project activities. This situation gives us the chance to try in the companies' context the standardized testing construction procedures initially created for the academic world.

Among the types of docimological services requested by the four companies and organizations, we mention:

- Training the teams in charge of the construction of standardized tests to the CCQCST model and its procedures for all four companies and organizations;
- Docimological analysis of the MCQs banks pre-existing in the companies and organizations with proposals for improvements. This kind of activity is proposed at the 3<sup>rd</sup> step entitled "QUESTIONS" of the CCQCST model (for AXA and IFA);
- Accompaniment and follow-up of the trainers in charge of the realization of standardized tests which attest the trainees' competencies (for IFA). These kind of activities are proposed at the 1<sup>st</sup> step "ANALYSIS", the 2<sup>nd</sup> step "DESIGN", the 3<sup>rd</sup> step "QUESTIONS", the 4<sup>th</sup> step "TEST" and the 5<sup>th</sup> step "CORRECTION" of the CCQCST model.
- Reliability analysis of the results obtained at standardized tests. Proposals for an improvement of the reliability on the basis of the improvement of the MCQs and their solutions (for FOREM). This kind of activity concerns the 6<sup>th</sup> step "CORRECTION" of the CCQCST model.

On the one hand, these collaborations permit life tests of our docimological procedures. On the other hand, they give us the opportunity to try and improve an offer of docimological sub-

contracting services as a private company which valorises research results would have done it.

### *Relevance of the CCQCST model in the partners' context*

As we have just said, as soon as we explained the details of the CCQCST model to the companies and organizations which are partners in this research, four of them requested immediate docimologic services. The tasks requested were in direct connection with the CCQCST model we use for the development of the e-C&QCST platform. This experience provided us information about the relevance of our model in contexts of companies. It appears that our model is well appreciated.

We also interviewed persons in charge of creating standardized tests in the ten participating companies and organizations. They were asked to describe the actions that they realized during the construction of their standardized tests. A series of complementary questions were asked in order to determine certain aspects of the process, such as: collaborators concerned, temporal dimension, importance of the quality of the tests, insertion of the tests in other processes...

The documents created within the framework of the subcontracting docimological services offered by the SMART and the result of the interviews were discussed within the research group in order to analyze eventual readjustments of specifications of our model. It comes out that the CCQCST model and its procedures which were initially realized on the basis of docimologic researches and from former experiments with higher education teachers is relevant and well accepted by our partners coming from the private companies and organizations world. The model requires only small readjustments that we presented for validation to the ten private companies or organizations at a workshop meeting. The principal readjustment consisted of a clarification of the roles of the standardized tests developers. The difficulty lays in the fact that this definition of the developers' roles had to be extended so as to be used in very different contexts and that their roles should be assigned to a series of collaborators of one person in charge for the development of the standardized test (the head of the project) within the e-C&QCST platform. In other words, a standardized test construction is more often considered as a team policy in the world of the companies than in the academic world.

### *Users' roles within the e-C&QCST platform*

We propose seven roles within the e-C&QCST platform. Some of them can be distributed to one or more people. The first role is just a technical support for the installation of the software but the six other roles are directly concerned with the development of the standardized tests. The seven roles are:

- Role 1: "INSTALLER". The installer installs the e-C&QCST software on a server and after the installation, defines an administrator of the platform.
- Role 2: "ADMINISTRATOR". The administrator encodes the persons in charge of the test construction.
- Role 3: "TEST SUPERVISOR". The test supervisor defines and encodes the parameters of the platform. He encodes the tests developers, the collaborators or the subcontractors for the tasks related to the construction of a standardized test. The test supervisor has a right of access to all information and can directly intervene on all the subcontracted actions. He can also define a person who must give his opinion about the quality of the productions and deliver a quality label if the level of quality is sufficient.

The test supervisor must also define the public of evaluated persons. In an academic context, the test supervisor is generally the teacher in charge of the assessment of his students.

- Role 4: “TEST DEVELOPER”. One or more tasks related to the construction of the standardized test for every step of the CCQCST model are distributed to one or more developers by the test supervisor. The test developer can be a collaborator inside the test supervisor’s company or institution.
- Role 5: “QUALITY SUPERVISOR”. The quality supervisor has a right of glance without direct intervention on the productions of which he examines the quality. The results of the examination of the quality of the developers’ productions are put into the system. The quality supervisor can subcontract all his tasks or a part of them. The results are accessible to the test supervisor who can decide to make this information known to the collaborators and/or to the subcontractors and/or to the persons who are evaluated.
  
- Role 6: “QUALITY CONTROLLER”. One or more tasks related to the attribution of a quality label of the evaluation can be attributed to one or more persons by the quality supervisor.
- Role 7: “ASSEESSEES”. They are the persons who are evaluated (i.e. students or trainees). The test supervisor can give them partial access to the information pertaining to some parts of the eight steps of the construction.

These hypotheses concerning the roles related to the framework of the e-C&QCST users (G framework) will be put to test within the experimentation of the first modules which is presently beginning.

### *The SAFES project*

The immediate docimologic interventions requested by four companies and organizations of this project constitute an element in favour of the hypothesis of the existence of a market for subcontracting services in connection with the use of the e-C&QCST platform, an hypothesis of the economical valorisation framework (E).

Currently, the Walloon area encourages the economic valorisation of results of researches by supporting the creation of spin-off companies. Being given the success of the subcontracting service requested by four companies and organizations of this project, the SMART team filed in a project within the framework of program FIRST SPIN-OFF of the DGTRE. Our project entitled “Smart Advices For Evaluation Systems (SAFES)” (Gilles & Tinnirello, 2005) was well ranked by an independent scientific committee and accepted so that we are currently starting a two years feasibility study concerning a project of a spin-off company which would propose its services in the realization of assessments and standardized tests. At the end the two years, the creation of this spin-off company should coincide with the finalisation of the e-C&QCST platform so that this tool could be used by the spin-off team.

The SAFES project will allow us:

- To create a method of audit of the systems of testing;
- To conceive the procedures to be implemented within the framework of a quality control mark of the standardized tests;
- To define a typology of services which will be proposed to the future customers of the spin-off company;
- To realize a market research based on the methodologies of desk research and of field research.

## First interfaces of the e-C&QCST platform

The general ergonomics of the platform is as follows:

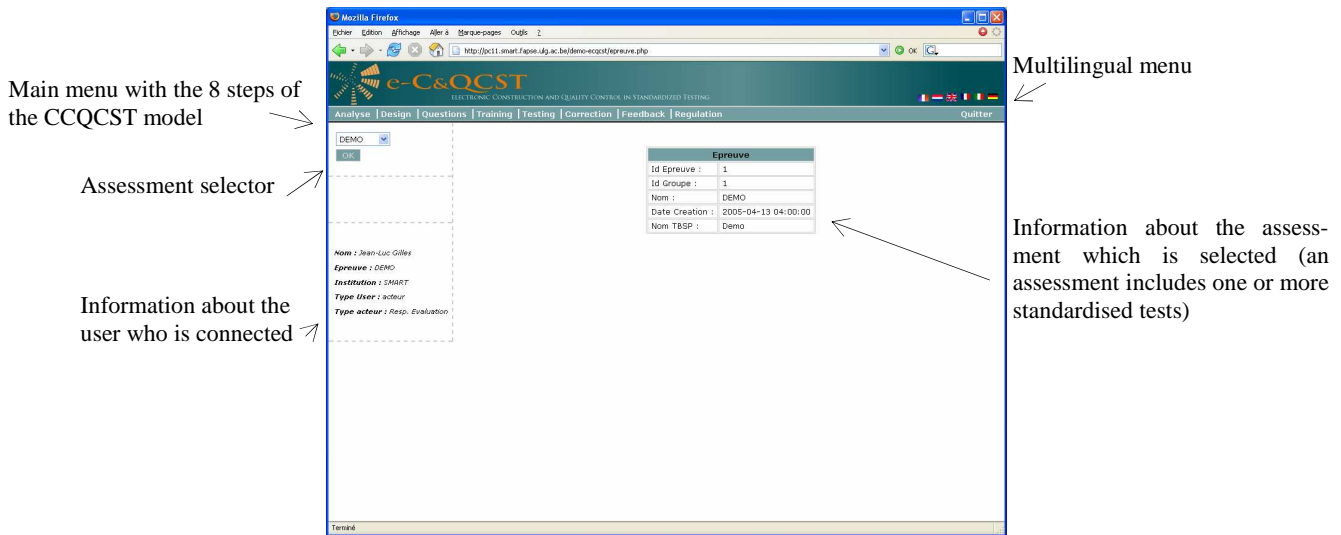


Figure 2. General ergonomics of the e-C&QCST platform.

We are actually testing with each of our partners the ergonomic choices we made for the interfaces. The first results concerning the intuitive utilisation of the platform are positive (hypothesis from the ergonomic framework [C]).

A positive effect of the meetings with the partners concerning the project specifications was to bring a clarification of the management of the collaborators and subcontractors intervening in the creation of the tests. We introduced the possibility for the test supervisor to create entries in the platform for various developers with reading or/and writing rights for the construction steps of a standardized test (hypothesis concerning the user's framework [F]).

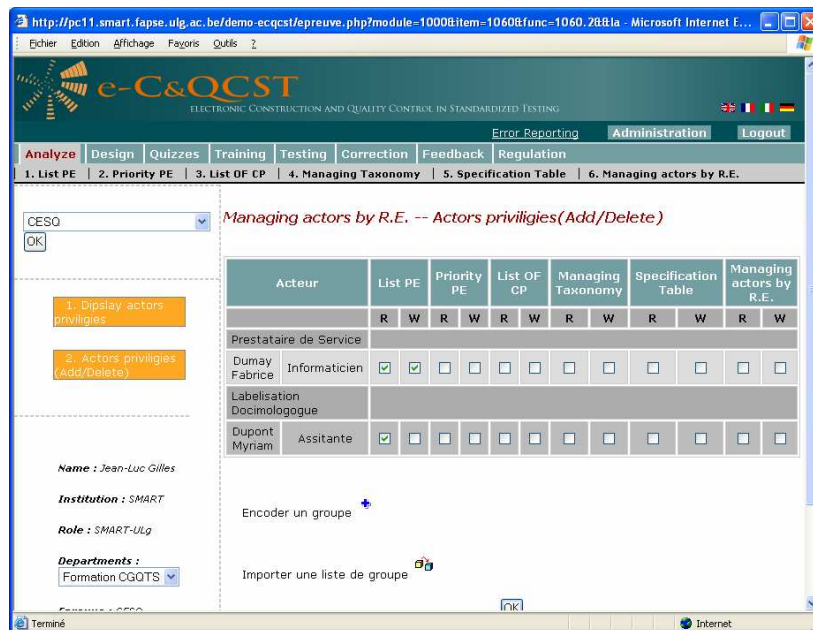
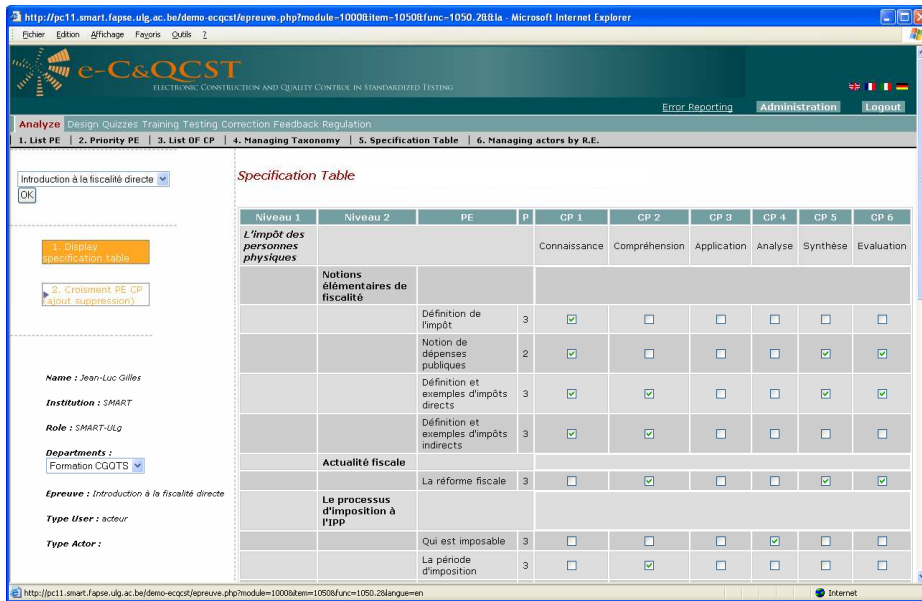


Figure 3. Management of the e-C&QCST users by the test supervisor.

Figure 2 shows the interface which enables the person in charge of the standardized test to define roles and rights in reading and writing within the framework of a series of tasks envisaged at the analysis module. In the example, the person in charge of the test defined a collaborator who is a member of the developers team and a data processing specialist (so called “informaticien” on the screen). The person in charge of the test granted him rights of reading and of writing for the task "List PE" which consists in drawing up the list of the points which are likely to be evaluated during the test (for example the contents list of a computer course). On the screen we can see that another person intervenes within the framework of the docimologic quality label. She must here control and check the quality of the work which is to be completed by the service provider in charge of creating the list of PE. She has only the right in reading and not in writing on the productions “List PE”. We mentioned this short example explicitly to fix the ideas about some possibilities to define the roles. Entering other persons in other categories of roles for other tasks is of course possible. Our hypothesis is that this kind of functionalities should contribute to a better coordination and visualisation of the standardized tests developers’ and user’s roles.

Actually, the first module entitled “ANALYSIS”, concerning the course/training or competencies description analysis, is finished. The productions of this module consist in a list of the points to evaluate (PE), of their priorities (P) and of the definitions of the categories of performances (CP) of the learners/trainees. The figure 4 presents an example of the end product which is represented by a test specification table where the points to evaluate and their priorities are crossed with the performances categories (PE x CP).



The screenshot shows a web browser window displaying the 'e-C&QCST' application. The main content area is titled 'Specification Table' and contains a table with columns for 'Niveau 1', 'Niveau 2', 'PE', 'P', and six 'CP' (CP 1 to CP 6) categories. The table lists various tax-related topics and their corresponding performance categories. Checkmarks in the CP columns indicate which categories are selected for evaluation.

Niveau 1	Niveau 2	PE	P	CP 1	CP 2	CP 3	CP 4	CP 5	CP 6
L'impôt des personnes physiques				Connaissance	Compréhension	Application	Analyse	Synthèse	Evaluation
	Notions élémentaires de fiscalité								
		Définition de l'impôt	3	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
		Notion de dépenses publiques	2	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
		Définition et exemples d'impôts directs	3	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
		Définition et exemples d'impôts indirects	3	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Actualité fiscale								
		La réforme fiscale	3	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
	Le processus d'imposition à l'IPP								
		Qui est imposable	3	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
		La période d'imposition	3	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Figure 4. End product of the 1<sup>st</sup> module ANALYSE: specification table where the points to be evaluated (PE) are crossed with the categories of performance (CP).

## d) Conclusions

We have defined six phases in the research and development planning of our 38 months project starting November 2003 and which is planned to end in December 2006:

- Phase 1: user's needs analyses, research frameworks, hypotheses and objectives definition, docimological procedures creation (months 1 to 8);
- Phase 2: conceptualization and preparatory project of each module of the e-C&QCST platform (months 5 to 17);
- Phase 3: writing of the project specifications for each of the 8 modules of the e-C&QCST platform, prototypes validation (month 8 to 32);
- Phase 4: experimentation in real size conditions of the platform modules (months 22 to 33);
- Phase 5: analyses of the experimentation results, improvement of the modules, final version creation and translation of the interfaces in the principal languages of the European Union (months 27 to 36);
- Phase 6: writing of the final report and spreading of the final results (months 35 to 38).

Actually we are in the middle of the 3<sup>rd</sup> phase of the project specifications and at the beginning of the 4<sup>th</sup> phase of the experimentation of the modules. The 1<sup>st</sup> module is being experimented with our research partners in academics' and companies' contexts.

The first lessons we can draw at this stage of the project, concerns the relevance of our CCQCST model, the users roles, the ergonomics of the platform and the economical valorisations perspectives of the results of this R&D project.

Concerning the relevance of the CCQCST model, we can say that it is well appreciated in the context of the ten companies and organisations which are our partners in this project.

Concerning the users' roles, the seven categories we defined were validated by our partners during a workshop. This clarification should contribute to highlight the complementarities and the responsibilities of the persons who are involved in standardized test creation.

Actually, we experiment the first module entitled "analysis" within the contexts of the ten companies and organization. The first results show that the ergonomics of the interfaces facilitate the production of specification tables related to the first step of the standardized tests construction.

Concerning the economical aspects, the research group is encouraged to create a spin-off company at the end of this project. The immediate docimological interventions requested by four of the ten companies and organisations partners in this project constitute an element in favour of the hypothesis of the existence of a market for subcontracting services in connection with the use of the e-C&QCST platform. The acceptance of the SAFES project in connection with these aspects gives us the possibility to make a feasibility study in this direction during the two next years.

\*\*\*

### e) References

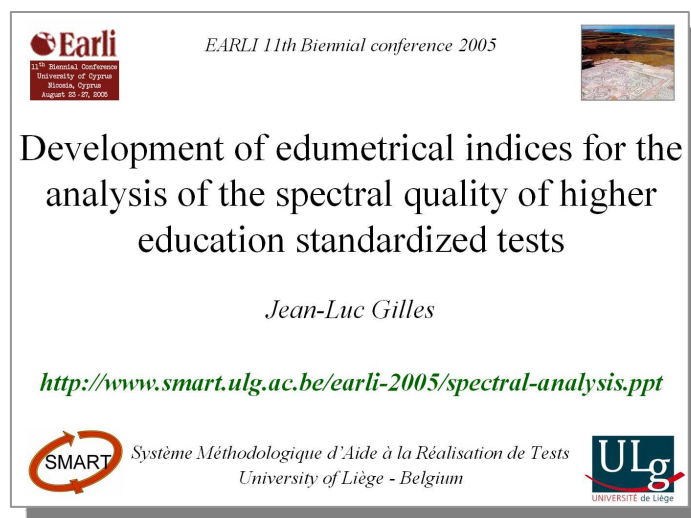
- Castaigne, J.-L., Gilles, J.-L. & Hansen, C. (2001). Application du cycle gestion qualité SMART des tests pédagogiques au cours d'Obstétrique et de Pathologie de la Reproduction des ruminants, équidés et porcins, Communication au 18ème Congrès de l'Association Internationale de Pédagogie Universitaire (AIPU), Les stratégies de réussite dans l'enseignement supérieur, Dakar, 5-7 avril 2001.
- Gilles, J.-L. & Leclercq, D. (1995). Procédures d'évaluation adaptées à des grands groupes d'étudiants universitaires - Enjeux et solutions pratiquées à la FAPSE-ULG, in Actes du Symposium International sur la Rénovation Didactique en Biologie. Tunis : Université de Tunis.
- Gilles, J.-L. (1998). *Introduction d'un système qualité dans le processus de réalisation des épreuves certificatives de l'enseignement supérieur*. Communication à la demi-journée de sensibilisation à la Qualité, 5 octobre 1998, Université de Liège
- Gilles, J.-L. (2002). *Qualité spectrale des tests standardisés universitaires*, thèse de doctorat en Sciences de l'Education. Liège : Université de Liège, Faculté de Psychologie et des Sciences de l'Education (559 pages).
- Gilles, J.-L. & Tinnirello, S. (2005). *Smart Advices For Evaluation Systems (SAFES)*, Proposition pour le programme "First spin-off" exercice 2005, Direction Générale des Technologies, de la Recherche et de l'Energie (DGTRE), Namur: Région Wallonne (76 pages).
- Kauffman, Ch., Dupont, P., Philippart, A. (1995). *Projet pilote européen pour l'évaluation de la qualité dans l'enseignement supérieur*. Bruxelles : Ministère de l'Education, de la Recherche et de la Formation.
- Kirkland, M. C. (1971). The effects of tests on students and schools. *Review of educational Research*, 41, 303-350.
- Kulik, C.-L., & Kulik, J. A. (1987). Mastery testing and student learning: A metanalysis. *Journal of Educational Technology Systems*, 15, 325-345.
- Leclercq, D. & Gilles, J.-L. (1993). Hypermedia : Teaching Through Assessment. In D. Leclercq & J. Bruno (Eds), NATO ASI Series, *Item Banking: Interactive Testing and Self Assessment*, Berlin: Springer Verlag, Vol. 112, pp. 31-47.
- Nightingale, P. & O'Neil, M. (1994). *Achieving quality learning in higher education*. London : Kogan Page.
- Phelps, R. (2005). *Defending Standardized Testing*, London: Lawrence Erlbaum Associates, Publishers.
- Ramsden, P. (1991). A performance indicator of teaching quality in higher education : the course experience questionnaire, *Studies in Higher Education*, 16, 129-150.
- Segers, M. & Dochy, F. (1996). Quality assurance in higher education : theoretical considerations and empirical evidences, *Studies in Educational Evaluation*, 1996, 22 (2), 115-137.
- Zink, K.J. & Schmidt, A. (1995). Measuring universities against the european quality award criteria. *Total quality management*, 6 (5-6), 547-561.

## 2. Development of edumetrical indices for the analysis of the spectral quality of higher education standardized tests

*Communication présentée par Jean-Luc GILLES lors du congrès international  
« EARLI 2005 – 11th Biennial Conference of the European Association for Research in  
Learning and Instruction (EARLI) »*

*Auteurs : Jean-Luc GILLES<sup>1</sup>*

[PowerPoint en annexe II]



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

## SPECTRAL POINT-BISERIAL CORRELATIONS FOR THE ANALYSIS OF THE ITEMS OF STANDARDIZED TESTS

Jean-Luc Gilles  
University of Liège, Belgium

### a) Introduction

The confidence degrees technique associated to multiple choice questions (MCQs) makes it possible to bypass the "binary" character of students assessment performance (the selected proposal is either correct or incorrect) provided that a series of methodological rules are followed called "admissible probability measurement procedures" by Shufford & al. (1966). Usually, the confidence percentages which accompany the MCQs answers are used to deliver more subtle feedback on each person's spectral performances. The innovative aspect of our approach lies in the fact that we have exploited the confidence percentages provided by the students to provide spectral information on the quality of the questions (as opposed to information on the quality of students' performances). Our research thus led to the development of a series of original indices for the analysis of assessments' spectral quality. These spectral indices are intended to be used when the assessor must highlight problematic MCQ and, within those, the proposals which contain anomalies.

**Keywords:** *Confidence degrees; standardized testing; point-biserial correlation*

By inviting the student to accompany his choice by the percentage of chances to be correct he attributes to his answer, we allow more subtleties in the analysis of his performances. The confidence scale adopted in this study offers 6 degrees of subtlety: 0%, 20%, 40%, 60%, 80% and 100% of confidence. At one extreme, the choice of a wrong answer accompanied by the maximal percentage of certainty (100%) presents the worst situation, in which the student provides an erroneous answer by estimating that he has a maximal chance to be correct. On the other hand, the student who answers correctly with a maximal certainty shows an assured performance. Between these two extremes, other zones can be identified in the "spectral analysis" of the performances. Jans & Leclercq (1997, 1999) propose a specific terminology. They distinguish "ignorance" (correct answer and weak certainty), from "partial knowledge" (correct answer and average certainty), and "perfect knowledge" (correct answer and high certainty). Such spectral gradations were also considered by these authors in the case of an incorrect answer ("*mistaken knowledge*" and "*fallacious knowledge*" or "*dangerous knowledge*").

Usually, the confidence percentages which accompany the MCQs answers are used to deliver more subtle feedback on each person's self-assessment performances (De Finetti, 1965; Echternacht, 1972; Leclercq, 1982; Bruno, 1993; Dirkzwager, 1993; Leclercq & Gilles, 1993; Bruttomesso & al, 2003; Gilles, 2004; Reach & al, 2005). The innovative aspect of our approach lies in the fact that we have exploited the confidence percentages provided by the students to provide spectral information on the quality of the questions (as opposed to information on the quality of students' performances). Our research thus led to the development of a series of original indices for the analysis of assessments' spectral quality. These spectral indices are intended to be used when the assessor must highlight problematic MCQs and, within those, the proposals which contain anomalies.

Our starting intuition for the construction of these new indices is as follows: logically the students who answer a question correctly should provide percentages of certainty higher than the students who answer incorrectly. Thus, for a multiple choice question which functions normally from the point of view of the certainty percentages use, we should observe among the students who choose the correct answer a tendency to answer with higher percentages of certainty and, in parallel, among the students who choose a wrong proposal a tendency to answer with lower percentages of certainty. We will then say that there is "spectral coherence". In the case where this situation does not arise, for example when students choose higher certainty percentages for one of the incorrect answers rather than for the correct answer, we are confronted with a problem of inconsistency in the use of the percentages of certainty, we will then speak of "spectral inconsistency".

To measure spectral coherence we created two new types of indices starting from the classical point-biserial correlation coefficient (classical  $r_{pb}$ ) calculation principle. The classical  $r_{pb}$  makes it possible to evaluate up to what extent each proposed alternatives solutions of each MCQ discriminates the students according to the criterion of the number of correct answers. Logically, one expects that the students who collect a high number of correct answers tend to choose the correct answer for a given question and that the students which collect a lower number have a tendency to choose an incorrect proposal.

The two new types of spectral coherence indices of measurement are: (1) the Spectral Contrasted  $r_{pb}$  ( $r_{pb} SC$ ) and (2) the Spectral Contrasted  $r_{pb}$  calculated after Turbo analysis ( $r_{pb} SCT$ ). During a former research (Gilles, 1998) we developed the  $r_{pb} S$  in order to analyze the tendency to use higher certainty in case of correct answers than in case of incorrect answers.

We use the name " $r_{pb} SC$ " to indicate the  $r_{pb} S$  is constructed with a "Contrasted treatment" which consists in the computation of the  $r_{pb} SC$  of an incorrect answer by using the data of the students who choose this wrong response in contrast with the data restricted to the students who choose the correct answer excluding the data from the students who choose an other incorrect answer. The advantage lies in the elimination of the data of the students who have chosen the other incorrect answers. This avoids introducing into the measurement of spectral coherence, the "background noise" generated by the data of the other incorrect answers.

The principle of "turbo analysis", consists in operating a selection in the data used for  $r_{pb} SC$  calculation on the basis of the level of realism reached by the students. We can thus increase the reliability of information related to the spectral indices by restricting the data to the students who make less errors in their self-assessments (in their use of certainty percentages). The name  $r_{pb} SCT$  indicates  $r_{pb} SC$  calculated with a Turbo analysis. The word "turbo" refers to the rise to power of the instrument in terms of the quality of information provided as one restricts oneself progressively to data from sets of students who make less and less errors in their self-assessments. We mention in the index's name the threshold of realism used to select the data. For instance  $r_{pb} SCT80$  was calculated starting from the data of the students whose realism is equal to or higher than 80 (those who make between 0% and 20% of errors in their self-assessments).

We tried to test these new spectral indices designed for detection of suspect alternatives within suspect MCQ using the data of several thousands of answers and certainty percentages collected during the ten standardized tests of the MOHICAN project (Leclercq & al, 2003).

They consisted in ten tests of knowledge of the principal subjects at the level of the end of secondary education which were submitted to groups of students entering first year studies in eight of the nine university institutions of the French Community of Belgium (the number of questioned students varied between 1.392 and 3.846 according to tests). These standardized tests consisted of MCQs for which students were invited to choose one answer and systematically accompany it with a percentage of certainty.

## b) Method

### *Item Bank*

In the early nineties, the inter-academic council of the French speaking Community of Belgium (CIUF) gave the mission to a research group coming out of 9 Belgian academic institutions to study the problematics of success or failure in the first grade of these institutions. This study issued a report called “*Franchir le cap des candis* (traduction: *to get over the first grade*)” (Leclercq & al, 1997). In their conclusions the members of this workgroup proposed the MOHICAN plan (*Monitoring Historique de Cohortes de CANDidatures universitaires*) whom aim was to gather reliable data on the knowledge of the students entering university and to furnish formative informations to students and teachers in the view of improving the chances to success. One year later the plan was launched by the Ministry of Education of the French speaking Community of Belgium.

The MOHICAN plan was constituted of ten different tests including a variable number of multiple-choice questions (MCQ). The subjects were the following (the number of questions appears between brackets): vocabulary (45), syntax and logical structuring of ideas (12), understanding of texts (6), reading of maps, graphics and tables in geography (10), mathematics (22), physics (10), chemistry (8) and biology (10), general knowledge in arts and culture (25), general knowledge in History and Socio-economy (25). These MCQs were created by experts coming out of the 9 universities of the French speaking Community, on the basis of their observations (in their institutions) on the difficulties experimented by the students of the first grade. For most of the tests these observations were supported by the results of students exams or inquiries addressed to university teachers.

### *Answering modalities*

The MCQs were composed of 3 to 5 propositions with most often two additional solutions “none” and “all”, so-called “general solutions” because they are systematically proposed for every questions of the test. One only proposition by MCQ might be called correct answer. The students had to accompany each of their answers by one of the six following confidence degrees: 0 (0%), 1 (20%), 2 (40%), 3 (60%), 4 (80%) and 5 (100%). The idea was to obtain more subtle results, as we have mentioned above, and to give the students a feedback on their realism, an aspect of their meta-cognition (are they able to evaluate themselves correctly? do they tend to underestimate themselves? or overestimate themselves?).

## Population

The academic institutions who participated in the MOHICAN plan were: the Catholic University of Louvain (UCL), the Free University of Brussels (ULB), the University of Liège (Ulg), The Academic Faculties Notre-Dame de la Paix in Namur, the University of Mons-Hainaut (UMH), the Catholic Academic Faculties of Mons (FUCAM), the Saint-Louis Academic Faculties in Brussels (FUSL), the Polytechnics Faculties of Mons (FPM) and the Agronomics Academic Faculties of Gembloux (FUSAG). Among these 9 universities, 8 accepted to submit the tests to their students entering the 1<sup>st</sup> year. The ULB only who was already running this type of test many years ago did not propose the MOHICAN tests to his students. The ten standardized tests were gathered in three categories according to the sections to which they were administrated: to all sections of the first year or only to some of them.

Table 1

*MOHICAN tests' population for students beginning studies in the faculties of agronomy, applied sciences, economy, law, medicine, philosophy, psychology, sciences and veterinary (UCL, ULG, UMH, FUCAM, FUSL, FPM and FUSAG)*

	<i>n</i>
Vocabulary test (VOC)	3.846
Syntax and logical structuring of ideas test (SYN)	3.739
Text comprehension test (COM)	3.420
Reading of maps, graphics and tables in geography test (GEO)	3.688

Table 2

*MOHICAN tests' population for students beginning studies in the faculties of agronomy, applied sciences, medicine, sciences and veterinary (UCL, ULG, FUNDP, FPM and FUSAG)*

	<i>n</i>
Mathematics test (MAT)	2.516
Physics test (PHY)	2.497
Chemistry test (CHE)	2.501
Biology test (BIO)	2.507

Table 3

*MOHICAN tests' population for students beginning studies in the faculties of economy, law, psychology and philosophy (UCL, ULG, UMH, FUCAM and FUSL)*

	<i>n</i>
General knowledge in arts and culture test (ART)	1.392
General knowledge in History and Socio-economy test (HIS)	1.410

## *Spectral Contrasted point-biserial correlation calculation for the correct solution ( $r_{pb} SC^C$ )*

The  $r_{pb} SC$  is calculated by correlating the confidence degrees (from 0 to 5 in the case of the MOHICAN research's data) which accompanied the answers to a MCQ with the choices/rejects (1 or 0) of the solutions of this MCQ. We propose the following adaptation of the formula of the traditional point-biserial correlation to calculate the Spectral Contrasted  $r_{pb}$  of a correct solution ( $r_{pb} SC^C$ ):

$$r_{pb} SC^c = \frac{Dx^c - Da^c}{\sigma} \sqrt{pq}$$

- with  $Dx^c$  = the average of the confidence Degrees of the students who chose the correct solution  
 $Da^c$  = the average of the confidence Degrees of the other students who did not choose the correct solution  
 $\sigma$  = the standard deviation of the confidence degrees of all the students  
 $p$  = the proportion of students who chose the correct solution of the question  
 $q$  = the proportion of students who did not choose the correct solution of the question ( $q = 1 - p$ )

For the readability of the formula we have added a c in exponent when one has to calculate the  $r_{pb} SC$  of a correct answer. In the case of an incorrect answer the formula is different. For the correct answer, the positive  $r_{pb} SC$ , generally obtained, shows to what extent those who have chosen the correct proposition have a greater confidence degree than those who failed.

### *Spectral Contrasted point-biserial correlation calculation for an incorrect solution ( $r_{pb} SC^I$ )*

When we calculate the  $r_{pb} SC$  of an incorrect proposition, we take in account the confidence degrees of the students who choose this distractor and the degrees of confidence of the students who had chosen the correct answer. We contrast the certainty related to the considered incorrect solution with the certainty related to the correct solution by excluding the certainty related to the other incorrect solutions. This is reason for which we called this treatment "contrasted". We propose to define the  $r_{pb} SC^I$  in the case of the incorrect proposals by:

$$r_{pb} SC^I = \frac{Dx^I - Da^I}{\sigma} \sqrt{pq}$$

- with  $Dx^I$  = the average of the confidence Degrees of the students who chose the incorrect solution for which we calculate the  $r_{pb} SC^I$   
 $Da^I$  = the average of the confidence Degrees of the other students who answered correctly  
 $\sigma$  = the standard deviation of the confidence degrees of the selected students, i.e. those which selected the incorrect answer considered and those which answered correctly  
 $p$  = among the selected students, the proportion of those which chose the incorrect answer for which we calculate the  $r_{pb} SC^I$   
 $q$  = among the selected students, the proportion of students who chose the correct proposal ( $q = 1 - p$ )

For an incorrect proposition, the correlation of the choices/rejects with the confidence degrees has, in principle, a negative result. Those negative  $r_{pb} SC^I$  results, which we were expecting, show to what extent the choices of an incorrect solution have, in the average, been accompanied by lower confidence degrees as compared with the confidence degrees used by those who have chosen a correct answer.

### *Turbo analysis principle and realism calculation for the $r_{pb} SCT$*

We propose to integrate the realism of the students in the calculation of the Spectral Contrasted  $r_{pb}$ . Taking the realism of the students into account in the  $r_{pb} SC$  permits to give more confidence to the values we have obtained. In effect, if we calculate the  $r_{pb} SC$  with the data coming out only of the students who estimate themselves particularly well, without not too many over or under estimations, we improve the reliability of the  $r_{pb} SC$  values. On the contrary, if we take in account all the data and if a good many of the students are not very realistic, our indexes are distorted, spoiled by the incoherent use of confidence degrees.

In that respect, we have calculated the  $rpbis SC$  selecting students with higher and higher realistic levels. The expression "Turbo analysis" used for this type of treatment is related with the improvement of the quality of the information gradually given by the instrument with the raising

levels of the "Turbo levels". This means that we progressively take into consideration the data of the students who commit less and less errors in estimating themselves.

The realism of a given student for a defined test is calculated on the whole set of confidences given by the student for this test. The formula used in the MOHICAN plan to calculate the Realism index ( $R_s$ ) of the subjects is the following:

$$R_s = 100 - AVAEC$$

with

$$AVAEC = \frac{\sum_i (|C_i - TE_i|) \cdot NU_i}{NR_t} = \text{the Absolute Value of the Average Error of Certainty}$$

$i$  = index of confidence degrees

$C_i$  = Value of the confidence degree  $i$  (in percent)

$NC_i$  = Number of correct answers for the confidence degree  $i$

$NU_i$  = number of uses of the confidence degree  $i$  (if  $NU_i = 0$  the index  $i$  is ignored)

$TE_i$  = Rate of correctness of the confidence degree  $i$  (in percent) =  $100 \times NC_i / NU_i$

$NR_t$  = Total number of answers for the test ( $\sum_i NU_i$ )

$NC$  = Total number of correct answers ( $\sum_i NC_i$ )

We propose the expression  $r_{pb} SCT$  followed by the indication of the  $R_s$  threshold (a number between 0 and 100) used to calculate the  $r_{pb} SC$  within the framework of a Turbo analysis.

We could for example calculate the  $r_{pb} SC$  starting from the data of these students whose  $R_s$  is equal to or higher than 70 with a step of realism of 10, until a higher  $R_s$  level of 100, which would involve the calculation of the  $r_{pb} SCT_{70}$ ,  $r_{pb} SCT_{80}$  and  $r_{pb} SCT_{90}$ . Within the framework of this process, the "Turbo analysis" name is to be put in relation to the possibility of connecting successive  $r_{pb} SCT$  treatments by requiring each time a higher level of realism for the selection of the data.

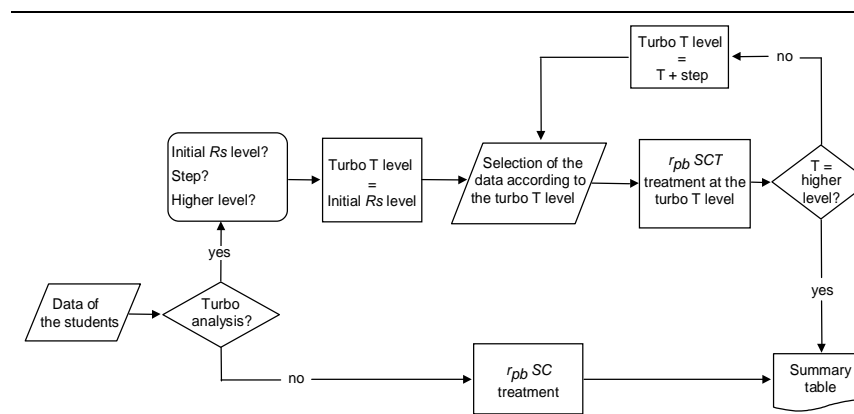


Figure 1. Turbo analysis process

**Null hypothesis for spectral contrasted correlations**

Contrary to the classical  $r_{pb}$  where the data of all the subjects are implied in the calculation of the correlations, all the couples of measurements are not used for the calculation of the  $r_{pb} SC$  and  $r_{pb} SCT$ , taking into account the contrasted treatment which is applied in the case of the incorrect solutions.

We showed above that the subjects implied in the calculation of the  $r_{pb} SC$  and  $r_{pb} SCT$  of an incorrect solution are on one hand those which had chose the incorrect considered proposal considered and on the other hand those had chosen the correct answer (from where the expression "contrasted" originates). This contrasted treatment eliminates the data of the subjects having chosen the other incorrect solutions and avoid introducing into the treatment the "noise" generated by the data of the other incorrect solutions. This generates also less couples of measures into the calculation of the  $r_{pb} SC$  and  $r_{pb} SCT$  of the incorrect solutions as in the case of the classical  $r_{pb}$ .

In addition, the point-biserial correlation being a Pearson's correlation, it is possible to test the null hypothesis  $H_0$  (the correlation is not significantly different from zero) in the same way as for  $r$ . In other words, the test of Fisher is applicable. It is possible to calculate a  $t$  which one may then be compared with the theoretical values of Student's table.

$$t = \frac{r_{pb} \sqrt{n-2}}{\sqrt{1-r_{pb}^2}}$$

with  $n - 2$  degrees of freedom.

In the case of the  $r_{pb} SC$  and  $r_{pb} SCT$  of an incorrect solution, being given the contrasted treatment,  $n$  is the number of couples of measurements for the considered solution added with the number of couples of measurements related to the correct answer, and not for the total number of couples of measurements as in the case of the classical  $r_{pb}$ .

c) Results

The indices of spectral quality can be calculated at various levels of turbo analysis. This makes it possible to increase the confidence in the information obtained since that at the highest levels we are only taking into account the data of the students who make less errors in their self-assessments.

However, the more we "go up" in the turbo analysis' levels, the less we have students. Consequently, it is important to know the expected number of students remaining at the successive levels of the turbo analysis.

In particular, are the population sizes at the T80 and T90 levels sufficient? We chose to fix the initial lower threshold of subjects Realism ( $R_s$ ) at 0, the  $R_s$ ' higher threshold at 100 and the step at 10.

The turbo analysis parameter setting involves 10 successive treatments. The first is made by taking into account all the subjects whose  $R_s$  lies between 0 (initial lower threshold) and 100 (higher threshold), in fact all the students.

Then, a second treatment is made by taking account of the data of the students whose realism lies between 10 (initial lower threshold increased with a step equal to 10) and 100 (higher threshold). This second treatment is followed by a third one where the data are those of the students whose realism lies between 20 (lower threshold of the preceding analysis increased with the step) and 100 (higher threshold).

The treatments follow one another, reaching each time a new turbo analysis level until the lower threshold is equal to the higher threshold minus ten. At the end, this parameter setting makes possible to calculate the spectral indices with 10 Turbo analysis' levels: T0, T10, T20, T30, T40, T50, T60, T70, T80 and T90.

Table 4

*Size of the 10 tests' population (VOC, SYN, GEO, HIS, ART, MAT, BIO, CHE and PHY) at 10 levels of turbo analysis (from T0 to T90 with a step of 10)*

	VOC	SYN	COM	GEO	HIS	ART	MAT	BIO	CHE	PHY
	<i>n</i>	<i>n</i>	<i>n</i>	<i>n</i>	<i>n</i>	<i>n</i>	<i>n</i>	<i>n</i>	<i>n</i>	<i>n</i>
<b>T0</b>	3.846	3.739	3.420	3.685	1.410	1.392	2.516	2.507	2.501	2.497
<b>T10</b>	3.846	3.734	3.413	3.683	1.409	1.392	2.514	2.507	2.500	2.494
<b>T20</b>	3.843	3.720	3.394	3.678	1.407	1.391	2.499	2.497	2.482	2.472
<b>T30</b>	3.838	3.699	3.348	3.649	1.405	1.387	2.465	2.485	2.461	2.458
<b>T40</b>	3.815	3.658	3.208	3.600	1.399	1.380	2.434	2.442	2.419	2.394
<b>T50</b>	3.773	3.553	2.860	3.474	1.391	1.363	2.398	2.340	2.341	2.342
<b>T60</b>	3.602	3.318	2.279	3.116	1.340	1.323	2.278	2.036	2.096	2.129
<b>T70</b>	3.089	2.691	1.481	2.303	1.101	1.137	1.956	1.511	1.661	1.724
<b>T80</b>	1.806	1.611	756	1.162	584	681	1.276	757	909	1.007
<b>T90</b>	350	439	251	253	73	124	339	151	232	275

We notice that with the T90 level the General knowledge in History and Socio-economy test (HIS) comprises no more than 73 subjects.

The other tests at this turbo level have all a population size higher than 150 subjects. With the turbo analysis level T80 all are high with  $n$  equal to 584 in the case of test "HIS". Among the 10 tests, it is that one which counts the smaller number of subjects at the turbo analysis level T80.

**Comparison of the  $r_{pb}$ ,  $r_{pb} SC$  and  $r_{pb} SCT80$  values obtained by a particularly discriminating question**

We calculated the values obtained by the 10 MOHICANs tests' questions to the point-biserial correlation coefficients calculated in a traditional way ( $r_{pb}$ ) and in a spectral way with contrasted treatment and turbo analysis ( $r_{pb} SC$  and  $r_{pb} SCT$ ). As regards to the  $r_{pb} SCT$ , we report here the values obtained at the turbo analysis T80 level, which seems to constitute a good compromise between the number of subjects available and the quality of the data. The ten MOHICAN tests counted on the whole 173 QCM. Among these questions, we located one of them, the 36th question of the vocabulary test (VOC Q36) which was particularly discriminating.

Table 5  
 Values obtained by the 36th question of the vocabulary test

	OM	P1	P2	P3
$n$	157	* 2.080	323	1.252
%	4	* 54	8	33
$r_{pb}$	-0.27	* 0.51	-0.16	-0.30
<b>VOC Q36</b> $r_{pb} SC$	xxx	* 0.41	-0.22	-0.35
$n T80$	103	* 1.199	135	612
% T80	5	* 58	7	30
$r_{pb} SCT80$	xxx	* 0.58	-0.28	-0.47

Note. OM = omission; P1, P2 and P3 = MCQ's solutions. Values obtained by the correct solution are preceded by a star (\*).

We propose not to calculate the  $r_{pb} SC$  et  $r_{pb} SCT80$  in the case of omissions because the students who omit should not, in theory, accompany their omissions by degrees of certainty, which explains the presence of the "xxx" in table 5. All the values of the  $r_{pb}$ ,  $r_{pb} SC$  and  $r_{pb} SCT$  correlations presented in table 5 are statistically significantly different from zero ( $p < 0.001$ ).

For this question VOC 36 we notice a positive  $r_{pb} SC$  for the correct response (0,41) and a negative one for the incorrect solutions P2 (-0.22) and P3 (-0.35). Concerning the  $r_{pb} SCT80$  we observe on the T80 level high population size for P1 ( $n = 1199$ ), P2 ( $n = 135$ ) and P3 ( $n = 612$ ).

The  $r_{pb} SCT80$  are even more marked than the  $r_{pb} SC$  for the correct answer P1 (0.58) and for the distractors P2 (-0.28) and P3 (-0.47). Let us recall that at the turbo analysis T80 level, calculations are obtained using the data of the students whose realism is very high ( $R_s \geq 80$ ).

We can conclude from this discussion that the use of the percentages of certainty is coherent: the students who answer correctly use higher confidence degrees than those who answer incorrectly. Thus, we observe a spectral coherence in the results of the question VOC 36.

With regard to the classical  $r_{pb}$ , let us recall that the choices or the rejections (1 or 0) of the MCQ's proposals by the students are correlated with the numbers of correct answers to all the questions in the test.

Logically, when the question functions in coherence with the other questions, one may expect that the choices of the correct solution are correlated with higher scores on the whole test and the choices of the distractors with lower scores. In this case, the question discriminates the students according to the criterion of the number of correct answers provided to the whole test. The subjects who collect a high number of correct answers tend to choose the correct proposal and the subjects who collect a lower number having the tendency to choose an incorrect proposal. We notice that the classical  $r_{pb}$  of the correct answer (0.51) is largely above the automatic correlation threshold ( $1/\sqrt{NQ} = 1/\sqrt{45} = 0.15$ ). As for the incorrect solutions, P2 collects a negative classical  $r_{pb}$  (-0.16) like P3 (-0.30). In short, the students who answer correctly thus tend to collect higher total scores and those who answer incorrectly lower total scores.

The classical  $r_{pb}$ , spectral  $r_{pb} SC$  and  $r_{pb} SCT80$  analysis agree with each other and make it possible to conclude there is no particular problem regarding to internal or spectral coherence for the 36<sup>th</sup> question of the vocabulary test.

*Comparison of the values obtained by  $r_{pb}$ ,  $r_{pb} SC$  and  $r_{pb} SCT80$  with the two most problematic questions of the MOHICAN tests*

The 3<sup>rd</sup> and the 20<sup>th</sup> questions of the *General knowledge in History and Socio-economy* test (HIS Q3 and HIS Q20) of the MOHICAN battery, collect abnormal values for the classical  $r_{pb}$  as well as for the spectral  $r_{pb} SC$  and the  $r_{pb} SCT$ .

Table 6

Values obtained by the 3<sup>rd</sup> and the 20<sup>th</sup> questions of the General knowledge in History and Socio-economy test

	OM	P1	P2	P3	P4	P5	P6
<b>HIS Q3</b>	<i>n</i>	90	501	132	197	104	258 * 128
	%	6	36	9	14	7	18 * 9
	<i>r pb</i>	-0.21	0.36	-0.02	-0.10	-0.06	-0.17 * 0.01
	<i>r pb SC</i>	xxx	0.49	0.05	0.07	0.04	0.19 * -0.15
	<i>n T80</i>	55	218	67	97	52	111 * 54
	% T80	8	33	10	15	8	17 * 8
	<i>r pb SCT80</i>	xxx	0.38	0.00	0.02	-0.03	0.06 * -0.06
	<b>HIS Q20</b>	<i>n</i>	58	263	39	* 256	20
%		4	19	3	* 18	1	0.5 53
<i>r pb</i>		-0.07	-0.18	-0.06	* 0.10	0.00	-0.09 0.14
<i>r pb SC</i>		xxx	-0.16	-0.10	* -0.17	0.03	-0.02 0.46
<i>n T80</i>		38	137	18	* 113	16	3 320
% T80		6	21	3	* 17	2	0 49
<i>r pb SCT80</i>		xxx	-0.24	-0.17	* -0.07	-0.05	-0.06 0.40

Note. OM = omission; P1 to P6 = MCQ's solutions. Values obtained by the correct solution are preceded by a star (\*).

Concerning the solutions of question HIS Q3, there is a problem of spectral coherence. The data of table 6 show that the *r pb SC* of the incorrect proposals P1 and P5 collect positive values (significance <.001), therefore the students tend to accompany these incorrect solutions by degrees of certainties which are in average higher than the certainties which accompany the correct answer whereas one expects the contrary. In addition, the incorrect solutions P2, P3 and P4 collect *r pb SC* which are equivalent to zero whereas one expects negative and more significant values.

Concerning the correct solution P6, we note that it poses also a problem given the negative value of *r pb SC* (significance <.001): one observes a weak correlation between the choices of smaller confidence degrees and the choices of this correct solution which should be, in theory, accompanied by the highest certainties (then should then have a significantly positive *r pb SC*).

When we consult the *r pb SCT80* (thus with the turbo level lower than T90) of P2, P3, P4, P5 and P6 we notice on the one hand minimum population size of 52 (P4) and on the other hand, quasi equal to zero *r pb SCT80*. Concerning P1, we notice that the *r pb SCT80* is positive: 0.38 (significance <.001).

The values obtained with *r pb SC* et *r pb SCT* thus indicate situations "of spectral inconsistency", the students tending to accompany the answer considered as correct with low confidence degrees and the incorrect solutions with higher ones. So, this question shows a problem of "spectral inconsistency".

Concerning the classical  $r_{pb}$ , this question HIS Q3 poses an internal coherence problem. We notice that the 1st solution (P1) collects a positive classical  $r_{pb}$  (0.36, significance < .001) whereas one expects it to be negative. Three other incorrect proposals: P2, P3 and P4, discriminate the subjects badly insofar as their classical  $r_{pb}$  are close to zero. There is also a problem with the correct answer P6 whose classical  $r_{pb}$  is quasi equal to zero (whereas one expects it to be positive and higher than the automatic correlation threshold value ( $1/\sqrt{NQ} = 1/\sqrt{25} = 0.2$ )).

With regard to question HIS Q20, the analysis of the propositions using on the one hand the  $r_{pb}$  SC and  $r_{pb}$  SCT80 and using on the other hand the classical  $r_{pb}$  lead to the same conclusion: for the two types of indices of discrimination (classical or spectral), the solutions P3 (correct answer) and P6 function in an incoherent way. We also don't notice classical or spectral discrimination for P4 and P5. With regard to P1, the two types of indices also converge by showing a coherent behavior.

On the other hand, for P2, we observe discrimination from the spectral point of view ( $r_{pb}$  SCT80 = -0.17, < .001) and no correlation with regard to traditional discrimination (classical  $r_{pb}$  = -0.06).

We thus notice a strong convergence of the problems highlighted by the various types of point-biserial correlations. When we ask for the opinion of the experts of the content, they confirm that these two MCQs are problematic. For the question HIS Q3 the distractor P1 could also be regarded as being correct and for other question HIS Q20, there is an error in the encoding of the correct solution (P6 instead of P3). For these two questions there is thus convergence of three different lightings: (1) that of the classical  $r_{pb}$ , (2) that of the experts and (3) that of the spectral coherence measured using the  $r_{pb}$  SC et  $r_{pb}$  SCT. In the case of the MOHICAN tests, the spectral analysis thus allows to highlight two questions that a more qualitative analysis (opinions of the experts) as well as a traditional analysis of discrimination (the classical  $r_{pb}$ ) also indicate to be questions with problems.

#### d) Conclusions

Let us point out the two principles which we applied to the Spectral point-biserial correlations.

On one hand we propose Contrasted treatment ( $r_{pb SC}$ ) for the incorrect proposals. It consists in contrasting the data of the students who choose an incorrect solution only with the data of the students who choose the correct solution. The advantage lies in the elimination of the data of the subjects who chose the other incorrect solutions. This avoids introducing into the spectral coherence measurement, the "noise" generated by the data of the other incorrect solutions.

On the other hand, we apply the principle of the "turbo analysis" to the spectral indices. It consists of operating a selection in the data used so as to keep for the calculation of the Spectral Contrasted  $r_{pb}$  only the results of the students who use the certainty in a judicious way, i.e. by making few errors in their self-assessments. We obtain more valid Spectral Contrasted Turbo point-biserial correlations ( $r_{pb SCT}$ ) because they are founded on the information provided by the subjects with more reliable data. Within the framework of the analysis we presented here, there are the  $r_{pb SCT80}$  calculated at T80 level, with the data of the subjects whose realism ( $R_s$ ) lies between 80 and 100 (maximum). In this study based on the MOHICAN tests data, T80 constitute a good compromise between the number of subjects available and the quality of the data.

The total number of MCQs for the ten MOHICAN tests was 173. For two MCQs among them, the 3rd and the 20th question of the *general knowledge test in History and Socio-Economy*, the values obtained with the spectral  $r_{pb}$  indicate situations of marked spectral inconsistency, the students tending to give lower percentages of certainty for the answers considered as correct and higher percentages of certainty for the incorrect answers. Studying the proposals of the two problematic MCQs by using the classical  $r_{pb}$  indices, we notice that the two MCQs do not function correctly from the point of view of classical discrimination. When we ask for the opinion of the experts of the contents, they confirm that the two MCQs display problematic results: for one of the questions a wrong proposal could also be regarded as being a correct answer and for the other, there is an error in the encoding of the correct answer. In the case of the MOHICAN tests, the spectral analysis thus allows to highlight two questions that a more qualitative analysis (posterior opinions of the experts) as well as a classical analysis of discrimination (classical  $r_{pb}$ ) indicate as questions with problems.

Does spectral analysis make it possible to do better than the classical analysis of discrimination for locating problematic MCQs and within them solutions which contain anomalies?

The answer must be refined. Within the framework of this study we analyzed the 173 MCQs of 10 MOHICAN tests by using two spectral indices  $r_{pb SC}$  and  $r_{pb SCT}$  as well as the classical  $r_{pb}$  indices. We also reviewed the comments the experts made about the contents of each question.

From these analysis, it appears that in addition to the two MCQs previously highlighted, 14 other questions present abnormal values for the classical  $r_{pb}$  and/or for the spectral  $r_{pb SC}$ .

What do the experts say? Among the 16 questions highlighted, they point five of them which pose real problems: the two MCQs we analyzed previously and the 14th MCQ of the *General knowledge in arts test*, the 6th of the *Biology test* and the 16th of the *General knowledge test in History and Socio-Economy*.

Concerning the point of view of the detection of real problems highlighted by the experts:

- The classical point biserial correlation ( $r_{pb}$ ) detects 2 problematic MCQs out of 5;
- The new spectral contrasted point biserial correlation ( $r_{pb SC}$ ) detects all the 5 MCQs highlighted by the experts;
- The new spectral contrasted point biserial correlation with turbo analysis at the level "T80" ( $r_{pb SCT80}$ ) detects 2 problematic MCQs out of 5.

Concerning the point of view of the "false alarms", in other words the detection of a problem whereas the experts do not detect a particular anomaly:

- The classical point biserial correlation ( $r_{pb}$ ) activates 10 false alarms;
- The new spectral contrasted point biserial correlation ( $r_{pb SC}$ ) activates 7 false alarms;
- The new spectral contrasted point biserial correlation with turbo analysis at the level "T80" ( $r_{pb SCT80}$ ) activates 1 false alarm.

Within the context of the MOHICAN tests, we can conclude, first, that from the point of view of "detection" of problematic MCQs, the new spectral contrasted point biserial correlation ( $r_{pb SC}$ ) is more effective in highlighting the problems raised by the experts. Second, that from the point of view of less "false alarms", the new spectral contrasted point biserial correlation with turbo analysis ( $r_{pb SCT80}$ ) is the more effective. From this point of view of less "false alarms" we can also say that without turbo analysis, the spectral contrasted point biserial correlation ( $r_{pb SC}$ ) is more effective than the classical point biserial correlation ( $r_{pb}$ ).

These qualities of better "detection" and less "false alarms" are crucial when highlighting the problematic MCQs is sought for.

When we correct the anomalies contained in certain questions, we can evaluate the impact of corrections on these proposals by comparing the values obtained before and after the operated changes. We did it and we observe noticeable improvements of spectral coherence.

Another point which should be underlined, is that the new spectral contrasted point biserial correlation ( $r_{pb SC}$ ) is "test free". With the classical point biserial correlation ( $r_{pb}$ ) it is necessary to take into account, on one hand, the choices and the rejects of the solutions of the MCQ and, on the other hand, the number of correct answers for the all test. So that, the quality of one MCQ is not independent of the quality of the other MCQs.

Now, with the new spectral contrasted point biserial correlation ( $r_{pb SC}$ ), on one hand we use the choices or the rejects of the solutions of the MCQ and, on the other hand, the confidence degrees given for that particular MCQ. So that we can say that the spectral contrasted point biserial correlation ( $r_{pb SC}$ ) do not depend of the quality of the other questions of the test, this index is "test free".

Using the spectral indices developed within the framework of this research, we open a new way for the analysis of the quality of the standardized tests and their regulation. From now on, we are able: to evaluate the spectral quality of standardized tests using confidence degrees; to highlight possible anomalies in the questions; and, after corrections, to evaluate the spectral impact of the improvements. This is our contribution to the improvement of procedures for high quality standardized tests, and, by extension, for the improvement of the reliability of the results.

### e) References

- Bruno, J. (1993). Using testing to provide feedback to support instruction: a reexamination of the role of assessment in educational organizations. In D. Leclercq et J. Bruno (Eds), *NATO ASI Series F, Item Banking: Interactive Testing and Self Assessment*. Berlin: Springer Verlag, Vol. 112, pp. 190-209.
- Bruttomesso, D., Gagnayre, R., Leclercq, D., Crazzolara, D., Busata, E., d'Ivernois, J.-F., Casiglia, E., Tiengo, A. And Baritussio, A. (2003). The use of degrees of certainty to evaluate knowledge, *Patient Education and Counseling*, Vol. 51, Issue 1, pp. 29-37, Diabetes Education and Quality of Life.
- De Finetti, B. (1965). Methods for discriminating levels of partial knowledge concerning a test item, *British Journal of Mathematical and Statistical Psychology*, 18, 87-123.
- Dirkzwager, A. (1993). A Computer Environment to Develop Valid and Realistic Predictions and Self-Assessment of Knowledge with Personal Probabilities. In: Leclercq, D., Bruno, J. Eds. *Item Banking: Interactive Testing and Self-Assessment*. NATO ASI Series F, vol. 112. Berlin: Springer Verlag p. 146-166.
- Gilles, J.-L. (1998). Apports des mesures métacognitives lors d'un test de compréhension d'un article scientifique, in Depover & Noël (Eds), *Approches plurielles de l'évaluation des processus cognitifs*. Mons : Université de Mons-Hainaut, pp. 19-30.
- Gilles, J.-L. (2002). Qualité spectrale des tests standardisés universitaires, thèse de doctorat en Sciences de l'Education. Liège : Université de Liège, Faculté de Psychologie et des Sciences de l'Education.
- Gilles, J.-L. (2004). Feedbacks diagnostiques via internet : un système individualisé de communication des niveaux de performances des étudiants dans le cadre de 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.
- Leclercq, D. (1982). Confidence marking, its use in testing, in B. Choppin & N. Postlethwaite (eds.), *Evaluation in Education : International Review Series*, Oxford: Pergamon, vol. 6, n°2, pp. 161-287.
- Leclercq, D. & Gilles, J.-L. (1993). Hypermedia : Teaching Through Assessment. In D. Leclercq et J. Bruno (Eds), *NATO ASI Series F, Item Banking: Interactive Testing and Self Assessment*. Berlin: Springer Verlag, Vol. 112, pp. 31-47.
- Leclercq, D., Peeters, R., Reggers, T., Charlier, J.E., De Ketele, J.M., Delhaxhe, M., Dupont, P., Lambert, J.P., Lambotte, J.P., Loeckx, E., Pilatte, A., Romainville, M. & Wolfs, J.-L. (1997). *Franchir le cap des candidatures – Inscriptions et réussites en candidatures universitaires dans la Communauté Française de Belgique de 1988 à 1995*. Bruxelles : Conseil Interuniversitaire de la Communauté Française.
- Jans, V. & Leclercq, D. (1997). Metacognitive realism: a cognitive style or a learning strategy?, vol. 17, nos. 1 and 2. *Educational psychology*, special issue of learning and strategies, p. 101-10.
- Jans, V. & Leclercq, D. (1999). Mesurer l'effet de l'apprentissage à l'aide de l'analyse spectrale des performances, in C. Depover & B. Noel (Ed.), *Evaluation des compétences et des processus cognitifs*. Bruxelles : De Boeck, pp. 303-317.
- Reach, G., Zerrouki, A., Leclercq, D. & d'Ivernois, J.-F. (2005). Adjusting insulin doses: from knowledge to decision, *Patient Education and Counseling*, Vol. 56, Issue 1, pp. 98-103.
- Shufford, A., Albert, A. & Massengil, N.E. (1966). Admissible probability measurement procedures, *Psychometrika*, 31, 125-145.

#### IV. *e-C&QCST* COMMUNITY : LES PRÉMISSSES D'UNE COMMUNAUTÉ D'UTILISATEURS DE LA PLATEFORME DE CONSTRUCTION ET DE GESTION QUALITÉ DES TESTS STANDARDISÉS

Prof. Dr. Jean-Luc GILLES,  
Coordinateur, chef de projet *e-C&QCST*

Pascal DETROZ,  
Chercheur *e-C&QCST*, Directeur du SMART (au 1<sup>er</sup> octobre 2005)

##### 1. Pertinence du concept de « Communauté d'utilisateurs »

###### a) Eviter de perpétuer avec un nouvel outil d'anciennes pratiques

Le but de ce projet *e-C&QCST* ne se limite pas à la seule livraison d'une plateforme en ligne de construction et de gestion qualité de tests standardisés, il s'agit aussi au travers des procédures proposées dans l'outil de contribuer à améliorer les pratiques docimologiques chez les enseignants et les formateurs... et ce n'est pas parce qu'on fournit un outil comportant des procédures de contrôle qualité docimologiques, qu'automatiquement la qualité des épreuves va s'améliorer. Le risque est grand de voir se reproduire avec de nouveaux outils d'anciennes pratiques. Dans le contexte de l'implémentation d'une plateforme d'enseignement à distance Claroline, Lebrun (2004) signale à de l'implémentation des innovations chez les enseignants : « Cette façon de procéder [utiliser un nouvel outil pour perpétuer d'anciennes pratiques] rassure en effet la personne qui entre en innovation. L'assimilation (utiliser l'outil nouveau pour faire comme avant précède bien souvent l'accommodation (imaginer de nouvelles niches, de nouveaux usages). Cette démarche d'assimilation – accommodation fait ressembler le processus d'innovation à un processus d'apprentissage. D'autres expérimentations et modèles (Concern-based adoption model, ACOT, Appreciative inquiry) nous confortent dans cette métaphore de l'innovation considérée comme processus d'apprentissage ».

Dans le contexte du projet *e-C&QCST*, nous pensons que le déclenchement d'un processus d'accommodation, d'apprentissage des procédures docimologiques que nous proposons à travers la plateforme pourrait être favorisé par l'insertion de l'utilisateur dans un réseau de personnes confrontées aux mêmes types de problèmes d'intégration de l'outil dans leurs pratiques. Une des fonctions du site communautaire que nous proposons en appui à la plateforme consisterait dès lors à favoriser les échanges, le partage entre utilisateurs en vue de mettre au point des usages pertinents d'*e-C&QCST* dans le monde de l'enseignement et de la formation en entreprise.

###### b) Favoriser une démarche de prise en main innovante

Pour favoriser une démarche de prise en main innovante de l'outil *e-C&QCST*, nous tablons sur le principe de « pluralité harmonique » (De Perretti, 1981) que nous résumons en trois mots-clés : diversification, collaboration et innovation.

*Diversifier* : c'est-à-dire envisager l'évaluation comme étant un processus multiforme où l'évaluateur souvent confronté à l'évaluation d'une large gamme de catégories de performances est amené à jouer la carte de la complémentarité entre différents outils d'évaluation.

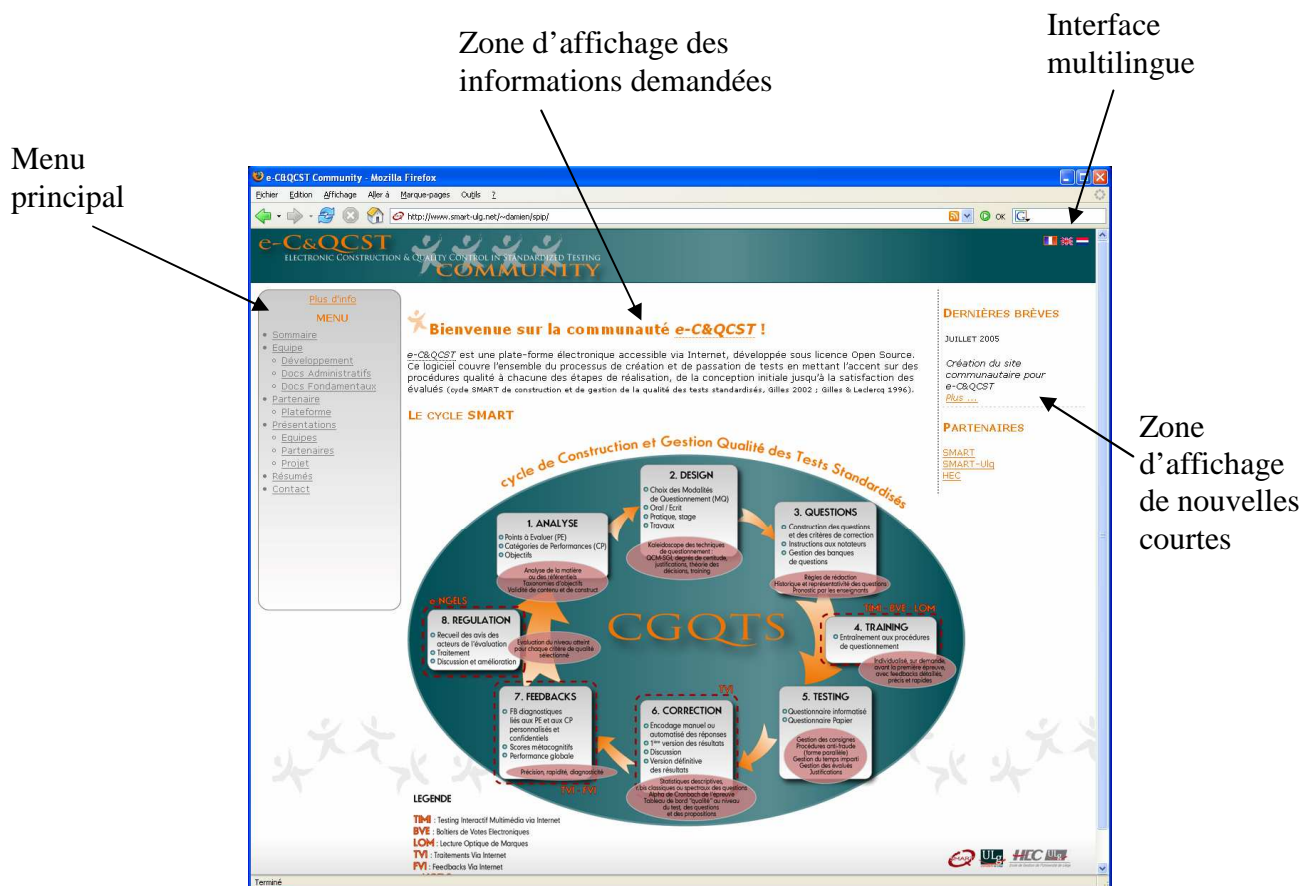
Il sera par exemple plus efficient d'évaluer le niveau « connaissance » de la taxonomie de Bloom à l'aide de QCM plutôt qu'avec des questions à réponse ouverte longue (QROL), mais par contre on ne pourra pas évaluer le niveau « synthèse » qui exige des productions avec des QCM, et, pour ce niveau taxonomique, les QROL restent un outil tout à fait pertinent, à condition toutefois de prendre des précautions pour contrôler la fidélité, l'équité, la validité, ...

**Collaborer** : amener l'utilisateur à travailler avec d'autres sur des projets d'évaluation, constitue selon nous un thème central à développer dans le cadre de notre « *e-C&QCST Community* ». Pour amener les utilisateurs à collaborer, dans une vision pluridisciplinaire, il faut organiser la confrontation des offres et des demandes de collaboration. Le site communautaire devrait jouer ce rôle en appui de la plateforme *e-C&QCST* qui, elle, gèrerait le travail collaboratif autour d'un projet d'épreuve à réaliser en commun.

**Innovater** : c'est-à-dire inventer du neuf en matière d'évaluation des apprentissages. La dynamique des échanges entre utilisateurs, le brassage d'idées, le travail collaboratif en équipes pluridisciplinaires, ... sont propices à l'émergence et au développement de nouveaux concepts docimologiques. Le site communautaire devrait, là aussi, stimuler la création et la diffusion de procédures d'évaluation innovantes.

## 2. Premières interfaces d' « *e-C&QCST Community* »

### a) Canevas de la page principale

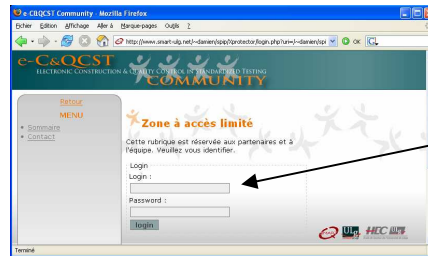


The screenshot shows the main page of the e-C&QCST Community website. Annotations point to various parts of the interface:

- Menu principal**: Points to the left-hand navigation menu.
- Zone d'affichage des informations demandées**: Points to the central content area containing the 'Bienvenue sur la communauté e-C&QCST!' message and the 'LE CYCLE SMART' diagram.
- Interface multilingue**: Points to the language selection icons in the top right corner.
- Zone d'affichage de nouvelles courtes**: Points to the 'DERNIÈRES BRÈVES' section on the right side of the page.

The central diagram, 'LE CYCLE SMART', illustrates the 'cycle de Construction et Gestion Qualité des Tests Standardisés' (CGQTS) with eight stages: 1. ANALYSE, 2. DESIGN, 3. QUESTIONS, 4. TRAINING, 5. TESTING, 6. CORRECTION, 7. FEEDBACKS, and 8. REGULATION.

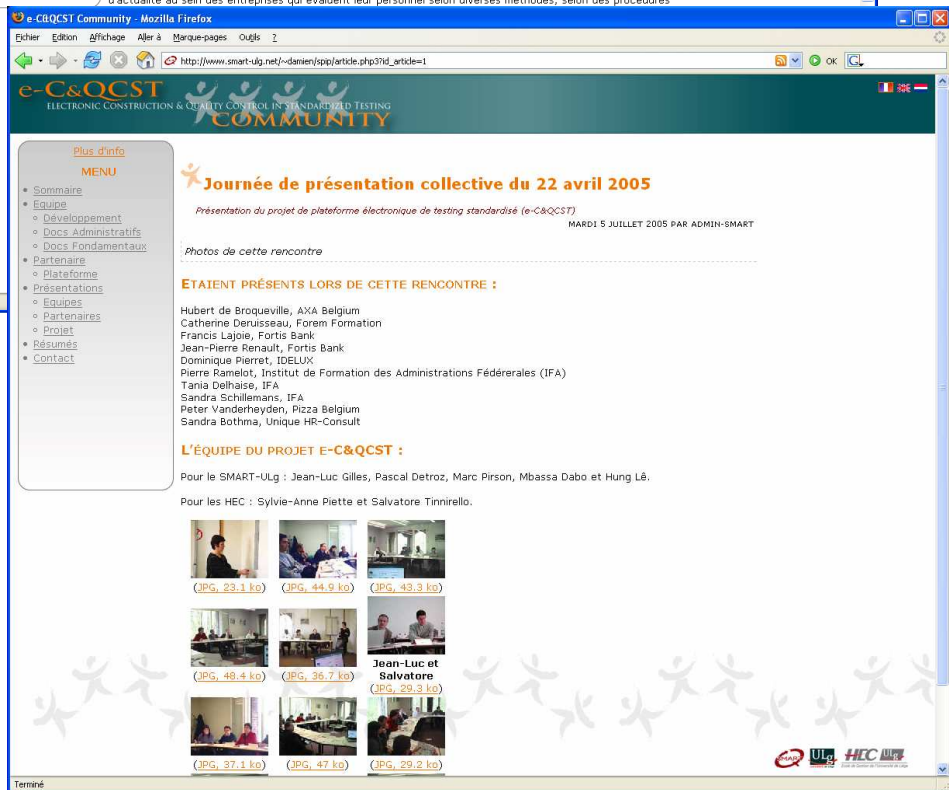
## b) Espace réservé à l'équipe



Accès limité via un « login » et un code d'accès

## c) Espace d'information publique

Présentation du projet via une des options du menu principal



### **3. Premières utilisations**

Les premières utilisations du site « *e-C&QCST Community* » auront lieu dans le cadre des expérimentations des modules 1 à 6 de la plateforme pendant le 1<sup>er</sup> semestre 2006 avec un groupe d'enseignants du monde de l'enseignement supérieur ainsi qu'avec un groupe de formateurs et d'évaluateurs provenant du monde de la formation en entreprises. Les demandes des partenaires de terrain dans le cadre de la phase d'expérimentation des modules devraient nous permettre de concevoir des fonctionnalités en adéquation avec les besoins réels de ces utilisateurs.

## V. ASPECTS SCIENTIFIQUES LIÉS À LA VALORISATION DE LA RECHERCHE *e-C&QCST*

### 1. Contextes de la valorisation économique des services de testing standardisés

*Prof. Dr. Jean-Luc GILLES,*  
*Coordinateur, promoteur du projet e-C&QCST*

*Salvatore TINNIRELLO*  
*Chercheur SMART porteur du projet FIRST SPIN OFF « Smart Advices For Evaluation Systems (SAFES) »*

#### a) Contexte socio-pédagogique

Nous assistons ces dernières années à une massification des effectifs dans l'enseignement supérieur organisé par des institutions publiques ou des organismes de formation. Selon Debry, Leclercq & Boxus (1998, p. 57), « ... *l'afflux d'étudiants dans le supérieur, et en particulier à l'université, pose notamment problème parce que les ressources d'encadrement n'ont pas suivi.* ». Dans la situation actuelle, la plupart des enseignants confrontés aux grands groupes ne disposent pas d'un nombre suffisant d'assistants pour permettre une correction objective et dans des délais raisonnables d'évaluations comportant de nombreuses questions ouvertes soumises à des centaines d'étudiants. Or, le sort académique ou la sanction d'une formation (réussite ou échec) et, par la suite, le sort professionnel de millions d'étudiants et de formés de par le monde dépend d'évaluations certificatives. On sait depuis des décennies (Pieron, 1963), à quel point la correction des examens où les étudiants sont interrogés à l'aide de questions à réponses ouvertes (écrites ou orales), souffre de nombreux biais. Ces biais expliquent en grande partie le succès rencontré par les épreuves standardisées dont le contenu et les conditions d'administration sont identiques pour tous les examinés et dont la correction peut être automatisée.

#### b) Contexte socio-politique

Dans le contexte économique actuel les entreprises démontrent une prise de conscience accrue de l'importance des investissements dans le capital humain : « *D'une part, elles sont confrontées à la mondialisation et à l'intensification de la concurrence ; D'autre part, le nombre de personnes âgées augmentant (également en leur sein !), les entreprises doivent apprendre à gérer le transfert progressif des compétences et, pour ce faire, les avoir préalablement identifiées, développées et valorisées.* » (Piette, 2002). Les entreprises et institutions de formation professionnelles se posent sans cesse la question de l'évaluation : évaluation des compétences pour assurer « un rendement » maximal au temps investi en formation ; évaluation en fin de formation afin d'identifier les acquis pour le travailleur mais aussi la qualité de la formation. Ces nouvelles compétences ne sont pas automatiquement suivies de performances sur le terrain puisque celles-ci sont également dépendantes des ressources et des objectifs liés à la fonction. Cependant, l'identification du niveau de compétences reste un indicateur très intéressant pour gérer la mobilité au sein de l'entreprise, l'accessibilité à certains projets, les niveaux d'exigences que l'on peut attendre d'une personne placée dans de bonnes conditions de travail. C'est par l'analyse transparente des perfor-

mances en regard des compétences qu'un travailleur (ou groupe de travailleurs) est en mesure de dialoguer sur les ressources qu'il a sa disposition ou qu'il devrait avoir pour atteindre les objectifs qui lui sont assignés.

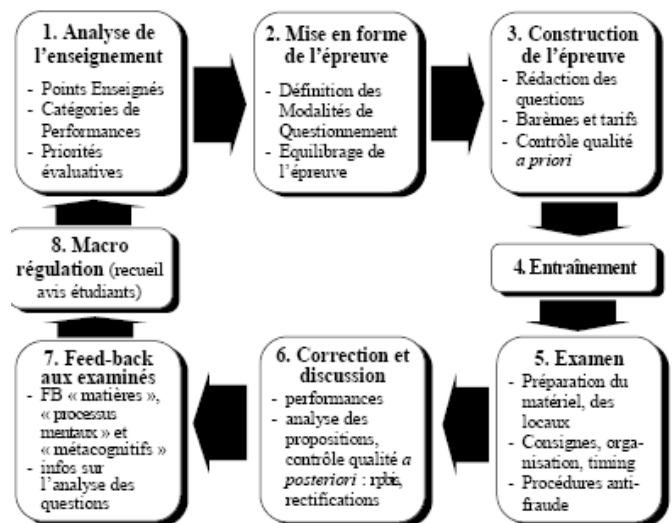
### c) Contexte scientifique

La recherche en docimologie et en éduométrie se doit d'offrir aux institutions d'enseignement supérieur / de formation des solutions ayant recours aux technologies de l'information et de la communication en vue de faciliter et de garantir la qualité des épreuves standardisées des acquis des apprenants.

Idéalement, ces solutions doivent s'inscrire dans un cycle en huit étapes de construction et de gestion de la qualité des évaluations (Gilles et Leclercq, 1995 ; Gilles, 2002).

Idéalement, ces solutions doivent s'inscrire dans un cycle en huit étapes de construction et de gestion de la qualité des évaluations (Gilles et Leclercq, 1995 ; Gilles, 2002). La figure ci-contre synthétise et schématise ces huit étapes.

Parallèlement, des garanties sérieuses de qualité ne pourront être offertes qu'à la condition d'envisager des procédures et des dispositifs d'ingénierie docimologiques qui rencontrent une série d'objectifs « qualité » :



- La *validité* : les scores des étudiants doivent refléter ce que l'enseignant/formateur veut mesurer ;
- 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 à un autre moment ;
- la *sensibilité* : la mesure doit être précise ;
- la *diagnosticité* : le diagnostic précis des difficultés d'apprentissage, des processus maîtrisés et de ceux qui ne le sont pas doit être possible ;
- la *praticabilité* : la faisabilité en termes de temps, de ressources en personnel et en matériel doit être assurée ;
- l'*équité* : tous les étudiants/formés doivent être traités de façon juste, en principe de la même manière (standardisation) ;
- 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.

Notre préoccupation relative à la qualité des évaluations standardisées s'insère dans mouvement mondial d'introduction du management de la qualité dans la sphère des activités de formation. Dès lors, une instrumentation de la gestion de la qualité des tests standardisés à l'aide du projet de plateforme *e-C&CQST* accessible via le web est susceptible d'intéresser un large marché européen d'institutions d'enseignement supérieur ou d'organismes de formation confrontés à la nécessité de

produire les preuves de la qualité de leurs tests certificatifs, notamment dans le contexte d'audits d'accréditation ou de la gestion des ressources humaines.

### d) Contexte économique

Le testing standardisé est une forme d'évaluation des compétences des apprenants qui, lorsqu'elle réunit les qualités de validité, fidélité, sensibilité, diagnosticité, équité, praticabilité, communicabilité et authenticité (Gilles & Leclercq, 1996 ; Castaigne, Gilles & Hansen, 2001 ; Gilles, 2002), est appréciée à la fois par les formateurs, les formés et les instances dirigeantes des institutions ou organismes de formation. Les préoccupations relatives à la qualité des évaluations standardisées des acquis des formés s'insère dans un mouvement qui depuis le début des années '90, touche les institutions belges et européennes (Kaufman & al., 1995 ; Segers & Dochy, 1996). Ce mouvement participe lui-même à une tendance mondiale d'introduction du management de la qualité dans la sphère des activités d'enseignement et de formation (Ramsden, 1991 ; Nightingale & O'Neil, 1994 ; Zink & Schmidt, 1995). Conscients de ces lignes de force actuelles, nous avons proposé en 2003 une recherche intitulée « *electronic Construction and Quality Control in Standardized Testing* (e-C&QCST) » visant la mise en place de méthodes et d'outils de gestion de la qualité des épreuves standardisées en vue de satisfaire au mieux les exigences des enseignants/formateurs et des étudiants/formés européens, voire d'autres continents. La dissémination et l'utilisation à l'échelle mondiale des résultats de cette recherche est possible à l'aide d'interfaces conçues pour fonctionner via le web. C'est l'idée du projet *e-C&QCST* de plateforme électronique de construction et de gestion qualité de tests standardisés réalisée à l'aide de logiciels libres et dans une philosophie « *Open Source* ». Dans ce contexte, après de nombreux contacts et investigations avec les partenaires du monde des entreprises et des grandes administrations/institutions publiques dans le cadre d'*e-C&QCST*, ainsi que dans le cadre des activités de prestations pour tiers du SMART, nous sommes amenés à faire les constats suivants : très forte demande de services dans le domaine de la construction et de la gestion qualité des épreuves standardisées, demande d'audit de la qualité de systèmes de testing, demande d'une forme de labellisation des évaluations certificatives (jusqu'ici non rencontrée).

## 2. Le projet First Spin Off “Smart Advices For Evaluation Systems (SAFES)”

*Prof. Dr. Jean-Luc GILLES,  
Coordinateur, promoteur du projet e-C&QCST*

*Salvatore TINNIRELLO  
Chercheur SMART porteur du projet FIRST SPIN OFF « Smart Advices For Evaluation Systems (SAFES) », Administrateur délégué de la spin off Assess-Group S.A.  
(au 1<sup>er</sup> octobre 2006)*

### a) Problématique de départ en lien avec e-C&QCST

Nous assistons depuis une décennie à un mouvement mondial d'introduction du management de la qualité dans la sphère des activités de l'enseignement supérieur (Ramsden, 1991 ; Nightingale & O'Neil, 1994 ; Zink & Schmidt, 1995). Les pratiques évaluatives dans les institutions d'enseignement supérieur n'échappent pas à cette tendance récente, c'est ce que soulignent Blais & al. (1997) : « *Dans le contexte social actuel, les établissements d'enseignement supérieur sont de plus en plus souvent amenés à faire la démonstration de la qualité de la formation dispensée. En ce sens, l'évaluation des apprentissages, comme mécanisme de régulation et de contrôle, exige une attention particulière. Par ailleurs, la régulation et le contrôle des apprentissages doivent prendre en considération un souci de plus en plus net d'assurer l'équité et la transparence des pratiques d'évaluation* ». Notre recherche se situe dans cette réflexion sur la mise en place de méthodes de gestion de la qualité des évaluations permettant aux institutions de faire la démonstration de la qualité de leurs épreuves certificatives. Plus particulièrement, nous nous attachons à la qualité des tests standardisés de plus en plus utilisés comme « *mécanisme de régulation et de contrôle* » dans des institutions qui doivent faire face à divers de facteurs de mutation depuis les années '60 dont l'augmentation constante du nombre de personnes à former et donc à évaluer et certifier.

Parmi les divers facteurs de mutation de l'enseignement supérieur, Dupont & Ossandon (1994, p.13) citent notamment la diversification de l'origine sociale des étudiants : « *... alors que les universitaires d'antan provenaient de filières secondaires sélectionnées, on observe que l'origine scolaire secondaire appartient à un éventail extrêmement varié de formations* ». A cette diversification sociale on peut aussi ajouter une tendance nouvelle de diversification des âges de ceux qui viennent se former à l'université. En effet, l'évolution rapide des sciences et des techniques entraîne la nécessité d'une formation continue. On remarque également la création de passerelles de plus en plus nombreuses entre les formations, favorisant ainsi encore plus de diversité dans les publics. Ces nouveaux publics posent aussi de nouveaux défis en matière d'équité dans la certification des compétences. Dès lors, rappelons que si dans un pays comme les Etats-Unis les tests standardisés connaissent un tel succès, c'est notamment parce qu'ils offrent des garanties de non discrimination raciale (*culture-free test*) et de non discrimination de classe sociale (*classfree test*).

Et à ce constat de la diversification des publics, s'ajoute celui de la massification. Récemment à l'Université de Liège, des amphithéâtres pouvant contenir plus de 600 étudiants ont été construits. Dès lors, la pédagogie des grands groupes s'impose par elle-même et comme le soulignent Dupont & Ossandon (1994, p.15) : « *par la force des choses, la pédagogie universitaire, prise sous l'angle des méthodologies utilisées, est de plus en plus inadéquate ; la plupart des universités européennes sont confrontées à une pédagogie des grands groupes à laquelle la majorité*

*des enseignants n'ont pas été sensibilisés ; on est loin d'un taux d'encadrement de 1 enseignant pour 15 à 20 élèves rencontré dans les prestigieuses universités américaines* ». Face à cette massification des effectifs, le testing standardisé s'impose comme une solution pragmatique pour la certification des compétences lorsque le contexte s'y prête. Par ailleurs, la pédagogie des grands groupes est souvent connotée grands auditoires. Notre projet de recherche se doit dès lors de mettre en place des stratégies qui permettent d'une part de garantir un haut niveau de qualité *a priori* des épreuves et d'autre part de faire la preuve *a posteriori* de la qualité des examens standardisés.

La problématique de la « qualité », historiquement née outre-atlantique dans les milieux industriels au début du siècle, pénètre progressivement monde de la formation. Au milieu des années 90, Barnabe (1995), puisant chez une série d'auteurs spécialisés dans le domaine de la qualité, faisait le pronostic suivant : « *Tôt ou tard, notre système d'éducation devra s'intéresser sérieusement à la qualité totale. Rhodes (1992) ainsi que Glasser (1992) et Deming (1992) prétendent que l'approche qualité totale est la seule réponse aux problèmes scolaires. Comme toute autre organisation, selon Kaufman (1992), un système d'éducation utilise des ressources, développe des moyens d'action et fournit des diplômés à l'attention des clients externes. Un fait demeure, un système d'éducation entretient la promesse qu'il produira quelque chose qui sera utile à la société qui paie dans l'espoir d'obtenir des diplômés de valeur. En dernier ressort, les commentaires reçus et les pressions exercées de la part des clients externes à l'égard de ces diplômés l'obligeront à rechercher la qualité totale* ».

Du point de vue de l'utilisation des TIC en Gestion des Compétences, la littérature et les colloques se multiplient sur le sujet tant il est d'actualité dans les entreprises et mêmes les institutions publiques (réforme COPERNICUS). En septembre 2001, l'ULg a organisé un colloque sur « la GRH dans et /ou la société de l'information » qui a montré l'actualité du questionnement méthodologique et sociologique des entreprises et des chercheurs face à l'utilisation des TIC. Cependant, la réflexion pédagogique concernant l'évaluation des compétences n'y a pas été abordée. Dans la littérature française concernant la validation des acquis de l'expérience ou dans la littérature provenant des entreprises (cadres, formateurs) est largement évoquée l'instrumentation de l'évaluation des compétences : les différentes méthodes d'évaluation et leurs avantages et inconvénients y sont listés sans pratiquement jamais mentionner et encore moins analyser l'utilisation de tests standardisés. Notre hypothèse est que ce type de test (avec ses composantes qualité) est assez méconnu et que cette recherche permettrait certainement d'aider à son développement en apportant les connaissances suffisantes pour montrer l'efficacité de ce dispositif.

### **b) Description du projet SAFES**

Le projet FIRST Spin off Région Wallonne SAFES consiste à valoriser une série de produits et de services issus de différentes recherches ayant trait à l'évaluation (au sens large du terme) dans les domaines de la formation en entreprise et des divers niveaux d'enseignements (scolaires et académiques). Les produits et services proposés couvrent des aspects aussi divers que l'évaluation de la qualité des formations ou des enseignements, l'évaluation des acquis des personnes en formation (à divers moments : à l'entrée, pendant, après...). Il peut s'agir d'évaluation des compétences, de testing standardisé (tests certificatifs, concours), d'évaluations formatives réalisées à l'aide de boîtiers interactifs, d'audit de la qualité de systèmes d'évaluations, de formations à la docimologie et aux méthodes dévaluation,... Un des produits provient notamment de la recherche e-C&QCST (programme WIST de la RW et de la DGTRE). Ce projet consiste à mettre au point une plateforme accessible via internet de construction et de gestion de la qualité des tests standardisés intitulée « *electronic Construction and Quality Control in Standardized Testing (e-C&QCST)* ».

L'idée est de permettre de créer suivant un modèle docimologique, des critères et des procédures qualité développés au sein du Système Méthodologique d'Aide à la Réalisation de Tests (SMART) de l'Université de Liège, des évaluations et pouvoir en maîtriser la qualité à différentes étapes. L'application a été créée en partenariat avec le monde de l'enseignement et des institutions publiques et privées (AXA, FORSTIS, IFA, FOREM, TNT, UNIQUE...).

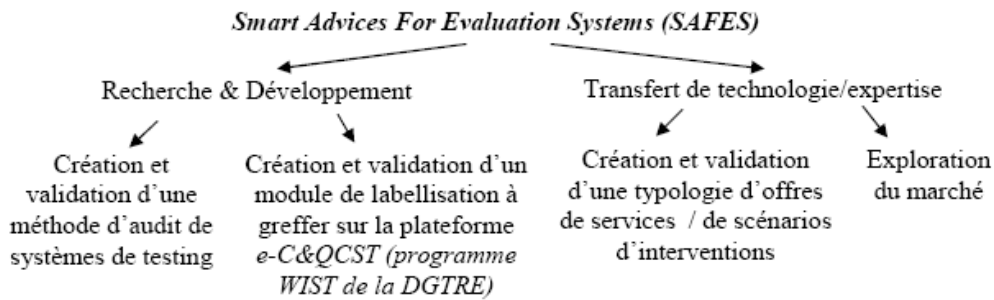
Cette application webbased suit les standards de programmation (serveur Apache, MySQL, PHP, normes W3C...). Un des buts de la spin off est de valoriser les services autour de cette plateforme e-C&QCST comme les formations, l'installation serveurs, développement de modules supplémentaires spécifiques, validations d'épreuves, coordination de concours à l'aide de la plateforme... Un des objectifs de la future spin off sera également de promouvoir l'idée de la création d'un label de qualité comme il en existe déjà un pour la certification de formation (QFor). A côté de ce produit, d'autres logiciels développés au sein du SMART pourront faire l'objet d'une valorisation, par exemple le logiciel TIMI (Testing Interactif Multimédia via Internet), le logiciel eNGELS (electronic Network for Global Evaluation of Learner's Satisfaction), etc. Pour certains, ils font déjà l'objet d'une valorisation ou sont en cours de valorisation au travers de GESVAL SA grâce à une série de prestations pour tiers que le SMART-ULg réalise déjà au quotidien.

### c) Objectifs poursuivis et approche méthodologique

Le recours à des méthodes et des dispositifs de testing utilisant les Technologies de l'Information et de la Communication (TIC) est incontournable pour répondre efficacement aux problèmes liés à la construction d'évaluations standardisées de qualité. La mise en œuvre d'une gestion qualité de ces tests s'appuyant sur les TIC est essentielle pour garantir un service optimal adapté aux exigences des enseignants/formateurs et des étudiants/formés, et, au-delà, des organismes de formation privés ou publics.

Par ailleurs, la pression sociétale est très forte sur les entreprises/administrations/institutions amenées à certifier les compétences des personnes qu'elles évaluent. En effet, les partenaires sociaux et les évalués eux-mêmes n'hésitent plus, le cas échéant, à introduire des recours en justice. L'équité des procédures est un sujet très sensible et remet à l'avant plan l'utilisation du testing standardisé. Il est donc crucial de proposer des méthodes d'audit, des outils de labellisation permettant d'apporter la preuve de la qualité des systèmes de testing. Nous pensons que de tels services ainsi que d'autres (à définir dans une typologie) pourraient être offerts par une spin-off active dans ce secteur.

Dès lors, les acquis scientifiques nouveaux attendus à l'issue de ce projet FIRST SPIN OFF relèvent d'une part de la Recherche et Développement et d'autre part du Transfert de technologie et d'expertise vers la société. Nous déclinons ces deux aspects en deux catégories d'objectifs comme présentés ci-dessous :



Le projet se structure autour d'un premier volet « Recherche & Développement » et d'un second volet « Transfert de technologie / expertise ». Au sein de chacun de ces volets, nous avons identifié deux lignes d'actions à mener en deux étapes. Chacune de ces lignes d'actions se décomposent en une série d'objectifs que nous détaillons ci-dessous.

**Étape 1 : Phase exploratoire**

**Volet Recherche & Développement**

*Ligne d'actions « Audit systèmes de testing »*

Objectifs	Méthodologie
<p><b>Conception de la méthode d'audit, des procédures, des documents, ...</b></p> <ul style="list-style-type: none"> <li>- Formalisation de normes à partir de nos modèles scientifiques</li> <li>- Définition des cadres d'application de ces normes</li> <li>- Identification des indicateurs des niveaux d'adéquation des normes aux contextes d'application</li> <li>- Définition des procédures permettant de vérifier les niveaux d'adéquation</li> <li>- Développement des supports pour l'audit</li> </ul>	<p>Ce travail de conception de la méthode d'audit sera effectué sur base d'une part de l'expertise en matière de testing standardisé présente au SMART et d'autre part de résultats de recherches déjà effectuées ou en cours de réalisation. Cette phase de conception aboutira à un avant-projet de manuel de procédures d'audit.</p>

*Ligne d'actions « Processus de labellisation »*

Objectifs	Méthodologie
<p><b>Conception des étapes du processus de labellisation, du label et du cahier des charges du module informatisé</b></p> <ul style="list-style-type: none"> <li>- Définition opérationnelle des critères et des seuils d'acceptabilité en lien avec les objectifs de qualité des tests</li> <li>- Elaboration d'un avant-projet de grille reprenant critères et seuils de manière à aboutir à des niveaux de label pour chaque étape du modèle</li> <li>- Construction du cahier des charges du module informatisé</li> <li>- Informatisé<sup>o</sup> du module labellisation</li> </ul>	<p>Utilisation des critères de qualité et du modèle de construction et de gestion qualité des tests standardisés (Gilles et Leclercq, 1996) utilisé dans le cadre de la plateforme e-C&amp;QCST. Cette phase aboutira à un projet de processus de labellisation formalisé dans un module informatisé relié à e-C&amp;QCST.</p>

**Volet Transfert de technologie/expertise**

*Ligne d'actions « Typologie d'offres / de scénarios d'intervention »*

Objectifs	Méthodologie
<p><b>Formaliser des lignes de produits et de services orientés vers l'évaluation</b></p> <ul style="list-style-type: none"> <li>- en support de la plateforme <i>e-C&amp;QCST</i></li> <li>- en support du module de labellisation</li> <li>- en support d'autres applications informatiques déjà développées au sein du SMART (<i>TIMI 2.0, e-NGELS, ...</i>)</li> <li>- définir les contenus exacts et les champs d'application de ces produits et services et leurs supports <i>ad hoc</i></li> </ul>	<p>Simulations d'implémentations de chacune des applications qui seront proposées en vue d'en déduire les services et produits nécessaires à leur mise en œuvre et leur appropriation effective.</p>

*Ligne d'actions « Exploration du marché »*

Objectifs	Méthodologie
<p><b>Etude de faisabilité</b></p> <ul style="list-style-type: none"> <li>- Réaliser une étude de marché sur les services et outils d'évaluation</li> <li>- Réaliser une étude de marché sur les normes, les agrégations, les certifications et labels</li> <li>- Définir un avant projet de stratégie marketing</li> </ul>	<p>Analyse du secteur belge de l'évaluation</p> <ul style="list-style-type: none"> <li>o logiciels existants</li> <li>o services déjà proposés</li> <li>o prix pratiqués</li> </ul> <p>Analyse du marché</p> <ul style="list-style-type: none"> <li>o segmentation du marché</li> <li>o cibles possibles</li> </ul> <p>Analyse des différentes formes des normes, certifications et de labels en matière de formation et d'enseignement et de la place de l'évaluation au sein de chaque norme, certification et label.</p> <p>Des stratégies de positionnement, de communication et de prix seront ébauchées au terme de ces deux études de marché ainsi que une première approche financière : simulation des coûts</p>

**Etape 2 : Phase opérationnelle**

**Volet Recherche & Développement**

*Ligne d'actions « Audit systèmes de testing »*

Objectifs	Méthodologie
<p><b>Validation de la méthode d'audit sur le terrain avec de futurs clients</b></p> <ul style="list-style-type: none"> <li>- Expérimentation des procédures consignées dans l'avant projet du manuel d'audit auprès des partenaires de terrain, futurs clients potentiels</li> <li>- Evaluation de la faisabilité de l'implémentation effective sur le terrain</li> <li>- Régulation et amélioration de la méthode (normes, procédures, indicateurs, supports)</li> <li>- Création du manuel dans sa version finale</li> </ul>	<p>Les procédures et supports figurant dans l'avant projet de manuel d'audit seront validés auprès des partenaires de terrain. Cette phase de validation aboutira à un manuel d'audit opérationnel.</p>

*Ligne d'actions « Processus de labellisation »*

Objectifs	Méthodologie
<p><b>Validation du module informatisé de labellisation auprès de futurs clients potentiels, régulation et amélioration des procédures</b></p> <ul style="list-style-type: none"> <li>- Présentation de la méthode de labellisation auprès de futurs clients de la spin-off</li> <li>- Labellisation expérimentale de tests auprès des clients potentiels</li> <li>- Essai du module informatisé de labellisation dans le cadre de l'expérimentation</li> <li>- Amélioration du processus et du programme informatisé de labellisation</li> </ul>	<p>Les partenaires de terrains de la recherche e-C&amp;QCST qui expérimenteront la plateforme dans leur contexte seront invités à entrer dans le processus de labellisation de leurs tests. Une version définitive du module de labellisation relié à la plateforme sera proposée à l'issue de cette phase opérationnelle.</p>

**Volet Transfert de technologie/expertise**

*Ligne d'actions « Typologie d'offres / de scénarios d'intervention »*

Objectifs	Méthodologie
<p><b>Validation auprès de futurs clients potentiels des lignes de produits/services formalisés lors de la 1<sup>ère</sup> phase exploratoire</b></p> <ul style="list-style-type: none"> <li>- Essais applicatifs chez des clients potentiels</li> <li>- Amélioration et affinements des produits/services</li> <li>- Mise au point des versions finales des fiches produits/services</li> <li>- Finalisation des manuels de procédures</li> </ul>	<p>Cette phase d'expérimentation en grandeur réelle se fera avec nos partenaires de terrain ainsi qu'avec les parrains industriels futurs clients potentiels de la spin-off.</p>

*Ligne d'actions « Exploration du marché »*

Objectifs	Méthodologie
<p><b>Approfondissement de l'étude de faisabilité</b></p> <ul style="list-style-type: none"> <li>- Etude de marché <i>field research</i></li> <li>- Etude financière (coûts de location, de sous-traitance, d'exploitation...)</li> <li>- Etablissement de la stratégie commerciale</li> <li>- Etude des aspects contractuels (contrats type, conditions de vente, licences d'utilisation, ...)</li> <li>- Etude des réglementations en matière d'opérateurs de formation (agrément pour les services de formation qui seront proposés)</li> <li>- Etude des réglementations en matière de certification et de labellisation</li> </ul> <p><b>Définition des conditions d'exploitation de la nouvelle spin-off</b></p> <ul style="list-style-type: none"> <li>- Etude de la localisation et de la structure nécessaire</li> </ul>	<p>Etablissement du <i>break even point</i> de chaque produit et service. Choix des canaux de communication, création des médias de communication.</p> <p>Consultation des organismes y afférant.</p>

### d) Innovation technologique apportée par le projet SAFES

Lorsqu'un enseignant ou un formateur est amené à créer une épreuve standardisée d'évaluation des acquis de ses apprenants il lui est actuellement difficile de faire la preuve que son test répond aux objectifs/critères de validité, de fidélité, de sensibilité, de diagnosticité, d'équité, de praticabilité, de communicabilité et d'authenticité que nous avons rappelés au point 5.1. A l'aide des procédures docimologiques et du logiciel accessible via le web que nous développons dans le cadre de la recherche *electronic Construction and Quality Control in Standardized Testing (e-C&QCST)* soutenue par la DGTRE dans le cadre de son programme Wallonie, Informatique, Société et Technologie (WIST), non seulement les enseignants/ formateurs mais aussi les étudiants/formés et les gestionnaires des institutions/organismes de formation bénéficieront des innovations technologiques suivantes.

#### Pour les enseignants/formateurs :

- Mise à disposition d'une plateforme de construction de tests standardisés via une démarche structurée et des procédures docimologiques permettant de maîtriser, de mesurer / contrôler et de rectifier les niveaux de qualité aux différentes étapes de réalisation des tests ;
- Gestion sécurisée et à distance des banques de questions ;
- Testing formatif ou certificatif de haute qualité avec possibilité de passation des tests à distance à l'aide de la plateforme ;
- Possibilité de couplage avec des dispositifs d'acquisition automatique de données (par exemple une chaîne de lecture optique de marques) dans le cadre de tests « papiercrayon ».

#### Pour les étudiants/formés :

- Mise à disposition via le web d'informations certifiées à propos des procédures de testing, ces informations étant gérées en temps réel par les équipes de formateurs/évaluateurs;
- Entraînement aux épreuves à l'aide de logiciels accessibles via le web et permettant d'obtenir des informations fiables, précises et diagnostiques sur l'état des compétences des étudiants/formés ;
- Plateforme intégrée de passation à distance des tests standardisés de certification ;
- Feedbacks de qualité (précision, exhaustivité, en temps réel, ...) après l'évaluation.

#### Pour les institutions/organismes de formation :

- Mise en place d'une plateforme intégrée permettant de contrôler la qualité des évaluations standardisées organisées au sein de l'institution/organisme ;
- Procédures docimologiques permettant de produire des preuves de la qualité des évaluations, ces preuves pouvant servir d'arguments en termes de marketing pédagogique ;
- Pilotage et régulation de la qualité des évaluations, notamment à l'aide du module de macro-régulation ;
- Réduction des coûts liés à l'évaluation des grands groupes.

En terme de valorisation par une spin-off des résultats de la recherche *e-C&QCST* via des prestations de services en rapport avec la plateforme, nos contacts avec les partenaires du monde des entreprises et des grandes administrations et institutions nous amènent à proposer dans le cadre de ce projet FIRST SPIN OFF, une recherche en vue de créer et de valider une **méthode d'audit** des systèmes de testing qui, en amont de l'utilisation de la plateforme, permettrait d'identifier les points faibles dans les procédures de testing proposées par les futurs clients de la spin-off.

**A partir des résultats de l'audit**, ces clients seraient orientés vers des offres d'interventions proposées par la spin-off et qui sont à définir dans une **typologie de services** qui exploiteraient les ressources offertes par la plateforme *e-C&QCST*.

**En aval de l'utilisation de la plateforme**, pour les clients de la spin off qui seraient amenés à devoir produire la preuve de la qualité de leurs tests (par exemple dans le cadre de tests certificatifs, de concours ...), un **module de labellisation** permettrait de certifier la qualité des épreuves via un label basé sur les résultats d'une série de contrôles qualité et une vérification de la conformité des procédures proposées. Cette démarche de recherche doit être soutenue par une exploration du marché en vue d'adapter au mieux les caractéristiques de la méthode d'audit, de la typologie de services et du module de labellisation aux besoins de ce marché. Dans ce contexte, le projet FIRST SPIN OFF, s'il était accepté par la DGTRE, constituerait une opportunité de préparer minutieusement le terrain en vue de la constitution, début 2007, d'une spin-off de services dans le domaine du testing standardisé de qualité.

Le premier tour d'horizon que nous avons pu effectuer dans le monde de la formation au sens large nous a amené à constater que l'offre des firmes actuellement actives sur le marché ne rencontre pas encore les besoins définis plus haut. **Notre spin-off avec sa plateforme électronique intégrée de construction et de gestion qualité des tests, sa méthode d'audit des systèmes de testing, son label qualité des tests et sa typologie de services et scénarios d'interventions offrirait une réponse globale et originale à une demande du marché qui n'est pas encore satisfaite.**

e) Calendrier de réalisation

		Recherche & Développement		Transfert de technologie / d'expertise					
		[1] Création et validation d'une méthode d'audit de systèmes de testing à utiliser en amont d'e-C&QCST	[2] Création et validation d'un module de labellisation à utiliser en aval de la plateforme e-C&QCST	[3] Création et validation d'une typologie d'offres de services / de scénarios d'interventions pour la future spin-off	[4] Exploration du marché et étude de faisabilité liée à la création de la future spin-off				
Phase exploratoire	1	Formalisation de normes, définition des cadres d'application, identification des indicateurs des niveaux d'adéquation des normes aux contextes	Conception grille de labellisation liée aux modèles SMART	Formaliser des lignes de produits et de services orientés vers l'évaluation en support de la plateforme, du module de labellisation, d'autres applications SMART	Réaliser une étude de marché sur les services et outils d'évaluation				
	2								
	3		Cahier des charges module informatisé de labellisation		Définir les contenus exacts et les champs d'application de ces produits et services et leurs supports <i>ad hoc</i>	Réaliser une étude de marché sur les normes, les agrégations, les certifications et labels			
	4								
	5		Programation du module informatisé de labellisation relié à la plateforme e-C&QCST			Définir un avant projet de stratégies marketing			
	6								
	7								
	8								
	9								
	10	Définition des procédures permettant de vérifier les niveaux d'adéquation, développement des supports pour la méthode d'audit							
	11								
	12								
Phase opérationnelle	13		Expérimentation des procédures de l'avant projet du manuel d'audit auprès des futurs clients potentiels, évaluation de la faisabilité de l'implémentation effective sur le terrain	Présentation de la méthode de labellisation auprès de futurs clients de la spin-off, labellisation expérimentale de tests, essai du module informatisé de labellisation dans le cadre de cette expérimentation			Validation des produits et services via des essais applicatifs chez des clients potentiels	Etude de marché <i>field research</i> , étude financière (coût de location, sous-traitance...), établissement de la stratégie commerciale, étude des aspects contractuels (contrats type, conditions de vente...), étude des réglementations en matière d'opérateurs de formation, en matière de certification et de labellisation	
	14								
	15	Régulation et amélioration de la méthode d'audit (normes, procédures, indicateurs, supports...), création du manuel d'audit dans sa version finale			Amélioration du processus et du programme informatisé de labellisation		Amélioration et affinements des produits/services		Etude de la localisation et de la structure nécessaire
	16								
	17								
	18								
	19								
	20								
	21								
	22								
	23								
	24								

### **3. De la recherche à la consultance : le choix stratégique d'un transfert d'activité opéré par l'équipe du Système Méthodologique d'Aide à la Réalisation de Tests (SMART) de l'Université de Liège**

*Prof. Dr. Jean-Luc GILLES,  
Coordinateur, chef de projet e-C&QCST*

*Pascal DETROZ,  
Chercheur e-C&QCST, Directeur du SMART (au 1<sup>er</sup> octobre 2005)*

*Salvatore TINNIRELLO,  
Chercheur e-C&QCST, Administrateur délégué de la spin off Assess-Group S.A.  
(au 1<sup>er</sup> octobre 2006)*

*Le texte qui suit a fait l'objet d'une communication lors du colloque organisé par le LENTIC « Intervenir dans le monde du travail : la responsabilité sociale d'un centre de recherche en sciences humaines » (30 novembre 2006)*

#### **a) Introduction**

Le Système Méthodologique d'Aide à la Réalisation de Tests (SMART) est une Unité de Soutien Logistique et de Recherche (USLR) spécialisée dans les domaines de la docimologie (la science qui étudie les comportements des examinateurs et des évalués, les systèmes de testing des acquis), de l'évaluation de la qualité des enseignements et des enquêtes à caractère pédagogique. Depuis sa reconnaissance officielle au sein de l'Université de Liège, en 1998, l'équipe du SMART répond à des demandes de consultance émanant du monde des entreprises et des administrations publiques. Ces prestations pour tiers prennent notamment la forme :

- d'accompagnements de formateurs dans la réalisation des évaluations de leurs apprenants ;
- d'aides à la mise en place de systèmes d'évaluation de la qualité des formations ;
- de supervisions et d'organisations d'opérations de collectes de données liées à la formation.

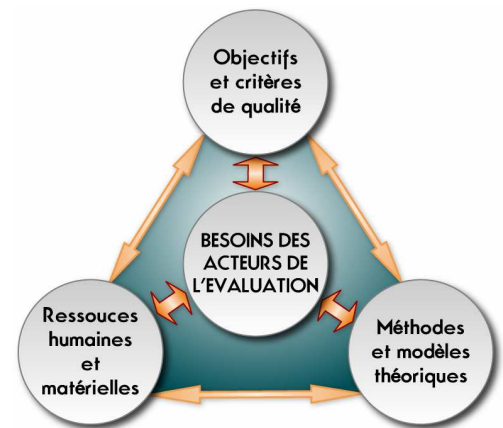
Depuis 2003, à l'occasion du projet de recherche et développement (R&D) e-C&QCST (Gilles & al., 2004, 2005 ; Crahay & al. 2006 ; Piette & al., 2006) de grande ampleur<sup>30</sup>, le SMART a multiplié ses contacts dans le monde des entreprises. Il en résulte une forte progression du nombre de prestations pour tiers. La demande est telle qu'aujourd'hui l'équipe composée d'une vingtaine de personnes (scientifiques, informaticiens et techniciens) s'est engagée avec l'aide de la Région Wallonne dans un projet de création d'une spin-off intitulé « *Smart Advices For Evaluation Systems (SAFES)* ». Il s'agit là d'un choix stratégique qui débouche sur le transfert des activités de

<sup>30</sup> Projet intitulé « *electronic Construction and Quality Control in Standardized Testing (e-C&QCST)* ». Programme WIST (2003) de la Direction Générale des Technologies, de la Recherche et de l'Energie (DGTRE).

consultance vers une entreprise externe qui devra assurer une valorisation commerciale des résultats de la recherche menée au sein du SMART. Les raisons de la création de cette spin-off sont épistémologiques, logistiques et pragmatiques.

## 1. L'approche qualité du SMART

Le SMART a développé une approche qualité (Gilles, 2002) qui soutient notamment nos prestations pour tiers. Cette approche qualité que nous défendons dans le cadre des activités de consultance repose sur quatre facteurs clés à prendre en compte dans la création d'évaluations. Notre approche est inspirée des recommandations de la norme internationale ISO 9004-2<sup>31</sup> (Gestion de la qualité et éléments de système qualité - Lignes directrices pour les services). Dans notre contexte lié à l'ingénierie docimologique, l'approche qualité proposée permet de focaliser les actions à entreprendre sur quatre facteurs clés. Au cœur de notre approche, le facteur central concerne les besoins des acteurs des évaluations, destinataires et utilisateurs des outils et dispositifs à concevoir, développer et appliquer. Ce facteur est aussi le point de convergence des trois autres : le facteur des modèles théoriques utilisés comme cadres de référence ; le facteur des ressources humaines et matérielles à mobiliser et le facteur des critères de qualité scientifiquement validés et qui nous guident dans nos actions. Notre approche qualité est utilisée depuis la fin des années nonante avec des partenaires du monde académique, du monde des entreprises et des grandes administrations. C'est dans cet esprit que le SMART a été lauréat d'un mandat de First spin-off de la Région Wallonne afin d'étudier la faisabilité d'une structure commerciale valorisant un ensemble de services et produits issus de la recherche en docimologie.



Cette approche pragmatique permet d'adapter de manière fine, sensible et pertinente les procédures, instruments et dispositifs scientifiques à mettre en œuvre dans les contextes de nos interventions.

## 2. Les missions du SMART

Le SMART est une USLR conçue pour gérer les demandes de l'Université de Liège (soutien logistique et méthodologique) et effectuer des recherches commanditées en rapport avec les trois missions suivantes :

- l'évaluation des acquis des apprenants : aider les enseignants à réaliser les évaluations de leurs étudiants (exercices, examens, tests formatifs...), notamment pour les grands groupes ;
- l'évaluation de la qualité des enseignements : aider les facultés à évaluer la qualité de leurs formations, à mener de grandes enquêtes de satisfaction auprès des étudiants ;
- les enquêtes à caractère pédagogique : aider les entités de l'ULg à réaliser des opérations de collecte de données en lien avec la formation des étudiants.

<sup>31</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.



Dans le domaine de l'évaluation standardisée des grands groupes d'étudiants, le SMART constitue une réponse de l'ULg au besoin d'évaluer rapidement et de façon fiable de grands auditoires. Grâce au SMART, les enseignants peuvent, par exemple, 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 :

- la consultation via Internet des feedbacks individualisés de leurs performances cognitives et métacognitives après une évaluation (un examen, une interrogation, une épreuve formative, ...)
- la réalisation d'entraînements aux procédures d'examen en amphithéâtre à l'aide de boîtiers de vote électronique ou sur Internet ;
- ...

Les recherches entreprises au sein du SMART ont permis d'articuler une série de critères de qualité docimologique avec les procédures d'un modèle original de construction et de gestion qualité des évaluations des acquis des étudiants (Gilles, 2002). Ces critères et ce modèle alimentent l'approche qualité dont nous avons parlé plus haut. Cette expertise nous permet d'accompagner des enseignants et des formateurs dans la réalisation de leurs épreuves certificatives et d'en labelliser la qualité.

En ce qui concerne ses ressources matérielles, le SMART gère une chaîne complète de lecture optique de marques, un dispositif de boîtiers de vote interactifs et des logiciels *ad hoc*. Parmi ceux-ci, le logiciel d'évaluation en ligne intitulé TIMI « *Testing Interactif Multimédias via Internet* » a atteint un niveau de maturité dans son développement informatique qui nous permet aujourd'hui d'en commercialiser des licences d'utilisation<sup>32</sup>.

Pour ce qui est de l'évaluation de la qualité des enseignements, le SMART fournit un support méthodologique et logistique à l'ensemble des facultés de l'ULg via la Commission interfacultaire « EVALENS ». Un logiciel spécifique à cette mission a été conçu et programmé par le SMART afin de permettre aux étudiants d'évaluer en ligne la qualité d'un enseignement et aux enseignants de consulter via Internet les synthèses des avis des étudiants. La solution offerte par ce logiciel intitulé e-NGELS « *electronic Network for Global Evaluation of Learners' Satisfaction* » (Detroz & al, 2006)<sup>33</sup>. Des licences d'utilisation de ce logiciel sont également commercialisées.

Enfin, en ce qui concerne les surveys, les dispositifs et les méthodes développés au sein du SMART permettent d'aider des personnes qui souhaiteraient réaliser des enquêtes dans le monde de la formation.

### b) L'afflux des demandes de prestations pour tiers

Progressivement, notre approche qualité dans le contexte de ce type de services d'abord fournis à l'échelle de la faculté de Psychologie et des sciences de l'éducation, puis, de plus en plus, à l'échelle de toutes les facultés de l'ULg, a commencé à intéresser le monde des entreprises et des grandes administrations. Sans prospection organisée, de plus en plus de prestations pour tiers ont

<sup>32</sup> Via la société Gesval qui valorise les résultats des recherches menées au sein de l'ULg.

<sup>33</sup> Cette méthodologie a fait récemment l'objet d'un reportage au journal télévisé de la chaîne française TF1 (reportage « Les professeurs notés par les étudiants en Belgique » - JT de 20 heures du lundi 26 juin 2006)

été effectuées. L'afflux de ces demandes a provoqué avec les recherches commanditées une augmentation graduelle de la taille de l'équipe. Constitué en 1998 de 3 personnes le staff est aujourd'hui passé à une vingtaine de membres (scientifiques, techniciens et secrétaires) dont près de la moitié travaillent sur des prestations pour tiers, les autres consacrant leur temps aux missions ULg ou aux R&D.

Dans le contexte de notre approche qualité, cet afflux des demandes de prestations pour tiers liées au testing des acquis, aux évaluations de la qualité des enseignements et aux surveys est somme toute assez logique car on assiste depuis le début des années nonante à une tendance mondiale d'introduction du management de la qualité dans la sphère des activités de formation. Cette tendance touche aussi les institutions belges et européennes (Ramsden, 1991; Nightingale & O'Neil, 1994 ; Zink & Schmidt, 1995).

### 1. Un contexte socio-économique propice

Voici quelques exemples qui montrent à quel point le contexte socio-économique peut être actuellement propice au développement de services docimologiques lorsque l'évaluation constitue un enjeu pour les personnes évaluées, notamment lorsque des avantages financiers sont en jeu lors de tests certificatifs par exemple. Ce type d'évaluation fait alors l'objet de toute la vigilance d'organes de régulation et de contrôle (les partenaires sociaux). Dès lors, les entreprises, administrations et institutions de formation sont de plus en plus sensibles à la nécessité d'apporter la preuve de la qualité (notamment en terme d'équité et donc bien souvent en ayant recours aux tests standardisés) de leurs épreuves certificatives.

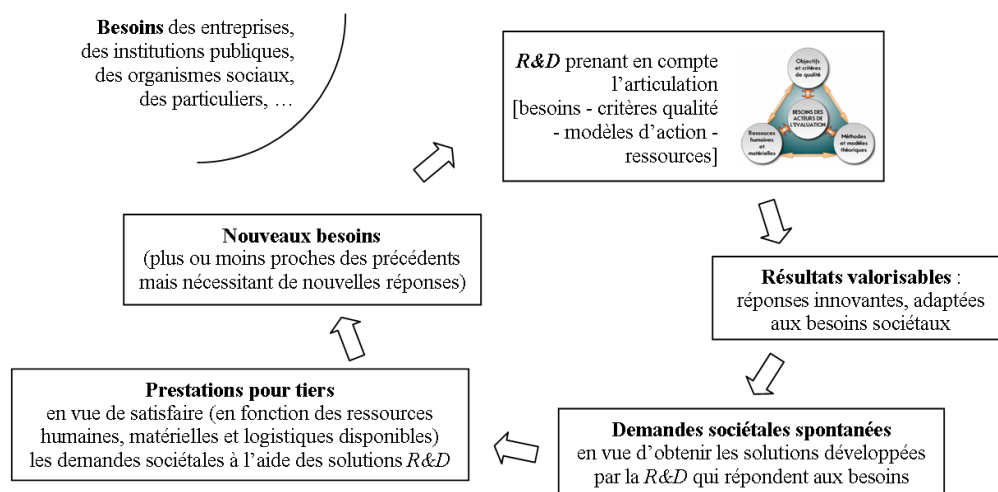
Il est admis que travailleurs et employés peuvent très bien développer dans leur travail ou en dehors de celui-ci, une série de compétences, souvent utilisées par l'entreprise mais non reconnues en dehors de celle-ci. Dès lors, « ...les individus en quête de reconnaissance durable de leurs acquis, cherchent à valoriser leurs compétences dans leur emploi ou, dans une optique de mobilité professionnelle (volontaire ou non) par le biais de reconnaissances officielles telles que certificats et diplômes. » (Piette & Orban, 2002). Fin 2001, le parlement a adopté une loi sur le "bilan de compétences". Cette loi, publiée au Moniteur le 30 décembre 2001, est suivie d'arrêtés d'application qui la rendent opérationnelle dès septembre 2002. Elle reconnaît à tous les travailleurs le droit à un bilan de compétences et le cas échéant à présenter des « examens » afin de faire reconnaître, valider et/ou certifier des compétences « *buissonnières* », c'est-à-dire acquises en dehors du cadre scolaire. Dans le contexte économique actuel les entreprises démontrent une prise de conscience accrue de l'importance des investissements dans le capital humain : « *D'une part, elles sont confrontées à la mondialisation et à l'intensification de la concurrence ; D'autre part, le nombre de personnes âgées augmentant (également en leur sein !), les entreprises doivent apprendre à gérer le transfert progressif des compétences et, pour ce faire, les avoir préalablement identifiées, développées et valorisées.* » (Piette, 2003).

Les entreprises et institutions de formation professionnelles se posent sans cesse la question de l'évaluation : évaluation des compétences pour assurer « un rendement » maximal au temps investi en formation ; évaluation en fin de formation afin d'identifier les acquis pour le travailleur mais aussi la qualité de la formation. Ces nouvelles compétences ne sont pas automatiquement suivies de performances sur le terrain puisque celles-ci sont également dépendantes des ressources et des objectifs liés à la fonction. Cependant, l'identification du niveau de compétences reste un indicateur très intéressant pour gérer la mobilité au sein de l'entreprise,

l'accessibilité à certains projets, les niveaux d'exigences que l'on peut attendre d'une personne placée dans de bonnes conditions de travail. C'est par l'analyse transparente des performances en regard des compétences qu'un travailleur (ou groupe de travailleurs) est en mesure de dialoguer sur les ressources qu'il a sa disposition ou qu'il devrait avoir pour atteindre les objectifs qui lui sont assignés.

## 2. Cycles R&D, prestations pour tiers et besoins sociétaux

Le schéma ci-dessous expose le type de cycle dans lequel s'inséraient les R&D et par la suite les interventions en entreprise du SMART dans les premières années de ses activités. Vu le contexte socio-économique (voir ci-dessus), les résultats de nos R&D ont très rapidement intéressé le monde des entreprises et des institutions, ce qui nous a amené à effectuer des prestations pour tiers. Celles-ci nous amènent à dialoguer de façon plus suivie avec nos partenaires de terrain, à nous insérer dans des réseaux et, au travers ces multiples contacts, à percevoir de nouveaux besoins qui peuvent entraîner de nouvelles R&D commanditées. Nous pensons que l'amorçage d'un tel cycle est amplement dû à la mise en œuvre de notre approche qualité qui place les besoins de nos partenaires au centre de nos interventions en les articulant avec des critères de qualité et des modèles d'action scientifiquement validés ainsi qu'avec des ressources efficaces et fiables. Comme nous allons l'exposer dans la partie suivante, il en découle une série de « *success stories* » qui au bout d'un temps peuvent amener une structure telle que le SMART à devenir la victime de ses succès.



## c) Gérer la valorisation des R&D dans le contexte du SMART

### 1. Vers une gestion de la masse critique

L'afflux des demandes de prestations pour tiers a eu différents effets sur le SMART :

- augmentation du nombre de personnes travaillant au sein de la structure ;
- diversification des profils de compétences au sein du staff (engagement de techniciens, de prestataires de services néerlandophones, d'infographistes, ... ) ;
- multiplication des contacts avec le monde des entreprises et des grandes administrations, les nouvelles demandes entraînant de nouveaux contacts et de nouvelles prestations pour tiers ;

- propositions de recherches commanditées en phase avec les besoins du terrain, ce qui augmente les probabilités d'acceptation de nos projets *R&D* ;
- ...

Comme nous allons le voir, ces effets positifs peuvent aussi engendrer des effets pervers.

### *Effets pervers d'une augmentation continue d'activités de prestations pour tiers*

Au bout d'un temps, l'afflux des demandes peut amener le cycle des R&D et prestations pour tiers à se gripper dans le contexte d'une structure universitaire telle que le SMART. En effet, l'augmentation et la diversification du personnel entraînent la nécessité de placer de nouveaux postes de travail dans des locaux qui ne sont pas extensibles. Dans notre cas des solutions ont d'abord été mises en place pour reconvertir des locaux inutilisés, comme par exemple un garage, mais l'accroissement du volume de demandes continuant à croître, de nouvelles personnes doivent être engagées et le problème des locaux resurgit...

Actuellement les demandes de prestations provenant de tiers sont spontanées sans qu'une « force de vente » consacre du temps à de la prospection de nouveaux clients. Il est évident que tôt ou tard la source des demandes spontanées se tarira peu ou prou et, sans une réelle politique commerciale et de marketing, l'équipe qui grossit actuellement sera tôt ou tard confrontée à un problème de masse critique. Actuellement les activités de prestations de services suivent une loi de Pareto, laissant poindre le danger du nombre restreint de clients. En effet, puisqu'un petit nombre de clients assure l'ensemble des prestations pour tiers, sans une augmentation significative (via une prospection active) du nombre de clients, la moindre diminution de contrats entraînerait des pertes d'emplois. Anticipant sur ce scénario probable, il est très difficile pour l'université d'accorder des contrats à durée indéterminée aux personnes engagées dans la structure pour travailler sur les contrats de prestations pour tiers. Cette situation entraîne un « *turn over* » important dans le personnel, ce qui engendre une perte d'expertise au sein du staff et complique encore la gestion des ressources humaines.

La multiplication des contacts avec nos partenaires de terrain est une bonne chose en ce qu'elle permet de mieux appréhender les réalités du terrain, mais avec l'augmentation du nombre de partenaires les responsables du SMART passent de plus en plus de temps à démêler des problèmes inévitables liés à la gestion des prestations pour tiers, ce qui les détourne de plus en plus de leurs missions au sein de l'université et de leurs recherches.

Alors que le SMART est en pleine phase de croissance en terme de prestations pour tiers, les conditions ainsi décrites pourraient être la cause à moyen terme d'une relative stagnation en terme de productions scientifiques et en terme d'émergence de nouveaux savoirs. Cela pose deux types de problèmes qui, s'ils ne sont pas prégnants actuellement, remettraient en cause l'existence même du SMART dans un avenir à moyen terme :

- Le SMART est une structure universitaire qui, comme toutes les structures universitaires, est jugée sur ses productions scientifiques. De plus, le SMART étant une USLR, il doit clairement continuer à effectuer des recherches qui bénéficient aussi à l'aspect logistique interne à l'institution universitaire, il s'agit de ne pas

gommer la *R&D* qui bénéficie à l'interne au profit d'une *R&D* uniquement orientée vers la réponse à des besoins externes ; un juste équilibre doit être trouvé ;

- Pour rester à un niveau de qualité optimale, les prestations pour tiers doivent pouvoir puiser dans des modèles théoriques en constante évolution et à jour scientifiquement. Ceci nécessite une veille scientifique large et donc en quelque sorte de « lever le nez du guidon » afin de maintenir à jour ses connaissances scientifiques. Ces moments de recul ne sont pas possibles lorsque les prestations pour tiers s'enchaînent les unes après les autres dans une structure qui n'est pas prévue pour gérer une activité commerciale sur un créneau très pointu.

### *Masse critique d'activités de prestations pour tiers*

La mise en œuvre de l'approche qualité définie plus haut à permis au SMART d'atteindre assez rapidement une masse critique d'activités de prestations pour tiers qui a placé la structure devant un choix : soit stopper l'augmentation des prestations et stabiliser leur volume de manière à mieux les maîtriser ; soit externaliser ces prestations dans une autre structure mieux armée pour gérer les demandes et apte à favoriser une croissance continue et rapide.

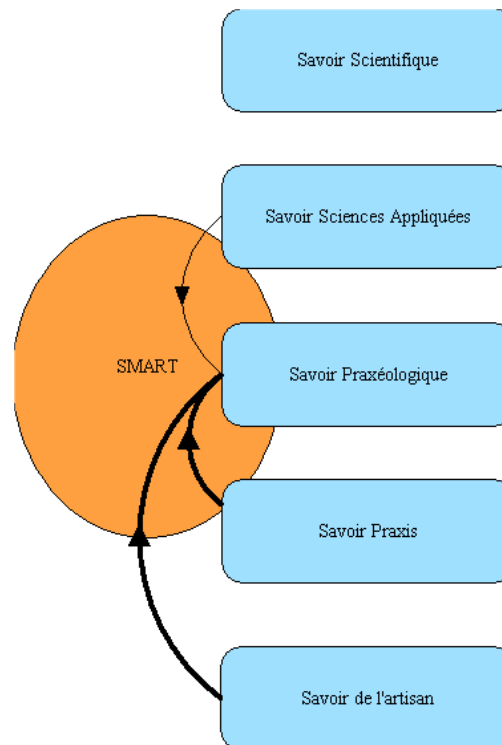
Selon nous, le premier scénario offre moins de potentialités que le second, mais à condition de veiller à maintenir un dialogue et des collaborations entre le SMART et la structure de valorisation qui diffuserait et exploiterait les résultats de nos *R&D*.

### *2. Vers une diversité épistémologique*

Notre approche qualité est utilisée depuis une dizaine d'années avec des partenaires du monde académique, du monde des entreprises et des grandes administrations.

Cette approche est pragmatique et permet de saisir pleinement le contexte dans lequel nos interventions prennent place. Elle permet d'adapter de manière fine et sensible les procédures et outils les plus pertinents à mettre en œuvre. Il s'agit, dans cette optique de proposer les outils les plus performants d'un point de vue scientifique dans le contexte analysé.

En ce sens, comme décrit dans le schéma ci-dessous, le SMART dans la taxonomie des savoirs de Van der Maren (1996) situe son action essentiellement au niveau praxéologique.



En maîtrisant parfaitement les modèles docimologiques, en possédant les compétences et les ressources humaines pour outiller ces modèles, et en inscrivant notre action dans l'atteinte de critères de qualité, nous sommes à même de fournir des solutions de qualité, adaptées et spécifiques à chacun de nos clients, qui s'adapteront, compléteront ou remplaceront leur pratique (praxis) en terme d'évaluation.

En terme de recherche, cette approche qualité est riche également car elle permet d'ancrer la recherche, les modèles et les outils créés dans la réalité des acteurs de l'évaluation.

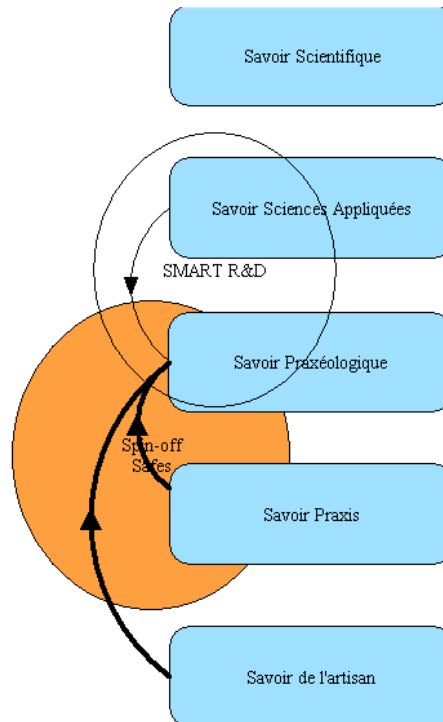
Le SMART est toutefois confronté à plusieurs difficultés :

- Les chercheurs du SMART travaillant sur les prestations pour tiers sont souvent pris par un jeu à double contrainte. D'une part ils s'inscrivent, en tant que prestataires de services, dans une relation commerciale avec une entreprise privée ou publique. Ils ont donc une obligation de résultats, et ces résultats doivent être, à tout le moins, satisfaisants pour l'entreprise. D'autre part, en tant que chercheurs, ils doivent, à plus long terme, produire des connaissances scientifiques. Il en résulte une négociation presque schizophrénique entre la posture de chercheur développant une vision à long terme et celle de praticien soumis à des contraintes de rendements commerciaux dans le court terme.
- Les solutions mises en place dans chacun des contextes de prestations pour tiers, même lorsqu'elles nécessitent une réflexion ou une démarche scientifique sont tellement spécifiques et contextualisées qu'il est difficile d'en tirer la quintessence en terme d'apport au corpus scientifique. Quand c'est le cas, ce savoir se situera le plus souvent au niveau praxéologique.

- Le SMART véhiculant une image de qualité universitaire, certains commanditaires sont, au départ, plus intéressés par l'aspect « validé par l'Ulg » que par la plus value procédurale apportée par notre expertise. Il existe un réel danger d'instrumentalisation du SMART dans un tel contexte.

Les dangers à moyens termes sont donc nombreux pour les chercheurs du SMART. En tant qu'unité de recherche, évaluée sur ses compétences scientifiques, il est important qu'elle puisse produire des connaissances d'un niveau supérieur au niveau praxéologique. De même, il est important pour ces chercheurs de développer une vision de recherche sur le long terme. Par ailleurs, pour les prestataires de la spin-off, se dégager de la posture de chercheur leur permettra d'être plus pragmatique et probablement plus efficace dans la mise en œuvre de solutions procédurales standardisées.

Ainsi, comme figuré dans le schéma ci-dessous, nous faisons le pari qu'une fois la spin off créée, le SMART pourra renforcer son expertise scientifique. De son côté la spin-off pourra monter en puissance en terme d'adéquation des solutions proposées aux contextes des clients.



### 3. Vers une rationalité des ressources

Dans le contexte actuel, les pouvoirs publics privilégient les programmes de recherches qui débouchent sur des résultats valorisables. Dans ce cadre, pour un centre de recherche universitaire comme le nôtre, avoir une synergie forte avec une entreprise axée sur la valorisation peut mener à un avantage concurrentiel par rapport à des centres similaires. De la même manière, une spin-off travaillant en lien étroit avec un laboratoire de recherche tirera un avantage certain de ce type de collaboration

#### d) Le scénario de la spin off « Assess Group Smart Advices For Evaluation Systems » (SAFES)

Selon nous, la structure de valorisation pourrait aussi jouer un rôle d'interface en vue de faire remonter les demandes sociétales vers le SMART. Par ailleurs, elle constituerait clairement un avantage pour le SMART dans le cadre de projets de recherches commanditées via des fonds régionaux et européens qui, de plus en plus, imposent des collaborations industrielles dans le cadre de la valorisation des *R&D* subventionnés.

Suite à ces constats nous avons déposé auprès de la Région Wallonne un projet de création d'une entreprise de valorisation de nos recherches dans le cadre du programme First spin off de la Direction Générale des Technologies, de la Recherche et de l'Energie (DGTRE). Le projet de spin off est intitulé « Smart Advices For Evaluation Systems » (SAFES).

A l'aide du financement de la Région Wallonne, le chercheur porteur du projet SAFES a pu, entre autres, effectuer une étude de marché. Celle-ci a mis en évidence la conjoncture actuelle très favorable pour une telle société, montrant à quel point d'une part l'évaluation des acquis de la formation est un domaine sous exploité commercialement en Europe (contrairement aux USA) et d'autre part, la qualité en formation constitue une préoccupation majeure (qui se concrétise au travers de diverses normes et de labels telles que EFQM, ISO<sup>34</sup>, Q\*For, EQUIS...).

Aidés par l'Interface-entreprises de l'Université de Liège, en moins d'un an un « *business plan* » a été réalisé, des licences de logiciels ont été mises au point, des actionnaires ont été réunis, des locaux ont été trouvés, une convention d'actionariat a été écrite, ... de manière à rendre la société opérationnelle dès septembre 2006 et à permettre un démarrage dans de bonnes conditions avec un staff de 5 personnes et un carnet de commande déjà bien rempli.

Lors de cette phase de préparation, l'équipe a pu prendre le temps nécessaire pour établir un *modus operandi* des collaborations et échanges ultérieurs entre le SMART et ASSESS GROUP. Les prestations pour tiers étant désormais assurées par la spin off, il fallait définir les mécanismes de retours vers le SMART. Ceux-ci seront assurés notamment via :

- la prise de participation de l'université de liège dans le capital de la spin-off ;
- la sous-traitance ponctuelle d'une expertise pointue du SMART dans le cadre de prestations effectuées par la spin off ;
- des collaborations sur les aspects *R&D* qui feraient partie de certaines prestations (par exemple des développements sur mesure à partir de produits ou services issus du SMART) ;
- la location de matériels et systèmes qui restent la propriété de l'ULg (par exemple la chaîne de lecture optique de marques, le système des boîtiers de vote électroniques, les serveurs web, ...)
- le dépôt de projets communs où des partenariats entre un centre de recherche et une ou plusieurs entreprises sont prérequis.

---

<sup>34</sup> Dans sa version 9001-2000, le chapitre 6 consacre une partie de sa norme à l'évaluation de la qualité de la formation.

D'un point de vue pragmatique, nous avons identifié la nécessité d'institutionnaliser un organe de communication et de discussion entre la spin off et le SMART. Il se concrétisera au travers d'un bureau scientifique composé du directeur du SMART, de l'administrateur délégué de la spin off et du responsable académique du SMART qui est aussi le promoteur du projet de spin off. Le danger serait pour le SMART et pour la spin off que ces deux structures ne communiquent pas. Pour le SMART, un dialogue collaboratif constant avec la spin-off est une nécessité pour garder un ancrage important avec les réalités du terrain. Pour la Spin-off, ce dialogue est le garant de la maîtrise continue des modèles et des procédures scientifiquement validées et mises à jour par SMART.

L'avenir nous dira si ce scénario tiendra ses promesses, mais on voit bien tout le potentiel qu'il recèle.

Nous pouvons d'ores et déjà entrevoir certains risques à ce changement de paradigme, notamment au niveau de la gestion des ressources humaines. En effet, des membres du personnel du SMART pourraient être tentés de quitter l'équipe de recherche pour rejoindre la spin off qui offre certains avantages matériels auxquels des chercheurs pourraient être sensibles. Une fois dans la spin-off ces mêmes chercheurs dans un environnement de travail différent pourraient à leur tour être nostalgiques de leur ancien métier de chercheur universitaire...

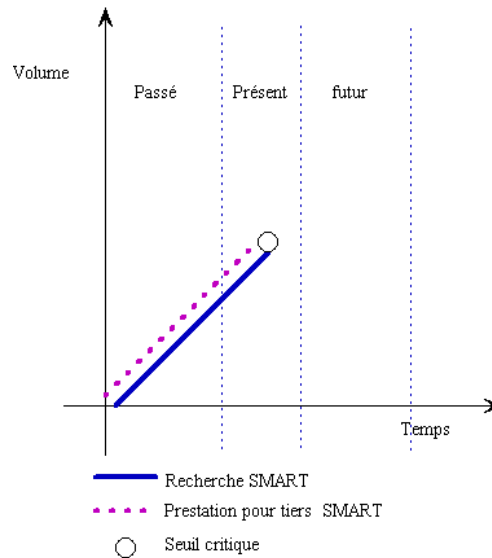
Par ailleurs, comment assurer un knowledge management efficace des nouvelles connaissances et compétences acquises par la recherche au sein du SMART, de tel manière à ce que la Spin off puisse maintenir un haut degré de compétences dans son domaine ?

A court terme nous entrevoyons des pistes de solution à mettre en place pour garantir cette gestion des compétences :

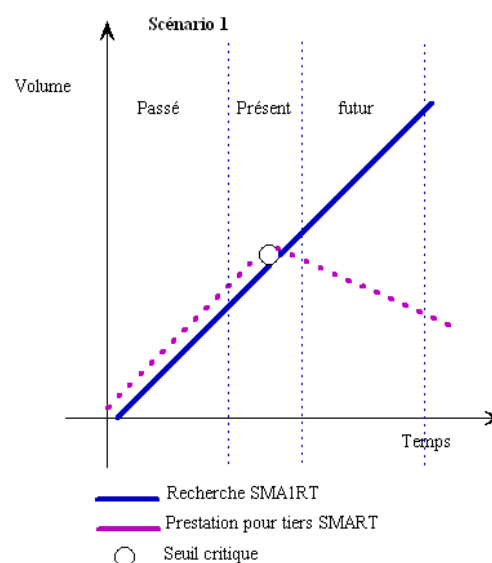
- mise en place d'un comité scientifique dans la Spin-off dont une des missions sera d'assurer un dialogue scientifique permanent avec l'équipe R&D du SMART
- organisation des collaborations scientifiques entre membres de la spin off et organes de l'université via la participation à des travaux pratiques, conférences, stages, ...
- participation à des projets R&D impliquant à la fois le SMART et Assess Group
- implication des membres de l'équipe du SMART dans des missions de service de la Spin-off
- Participation à des colloques et évènements scientifiques via des communications conjointes SMART-Assess Group

e) Conclusions

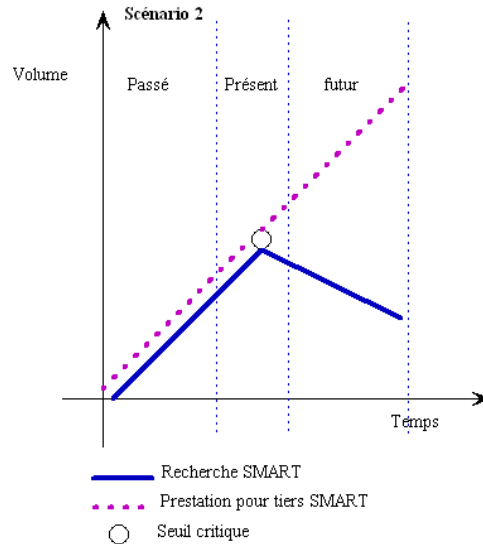
Le SMART a connu une phase de croissance rapide depuis sa création. Cette phase de croissance s'est réalisée sur deux axes : l'axe des prestations pour tiers et l'axe de la recherche.



Pour des raisons logistiques, épistémologiques ou pragmatiques précisées dans l'article, ces deux phases de croissances conjointes ne pouvaient continuer à être concomitantes à moyen et long terme. D'après nous, le SMART aurait été confronté à un choix difficile. Soit augmenter son volume de recherche en renonçant peu à peu aux prestations pour tiers (scénario 1).

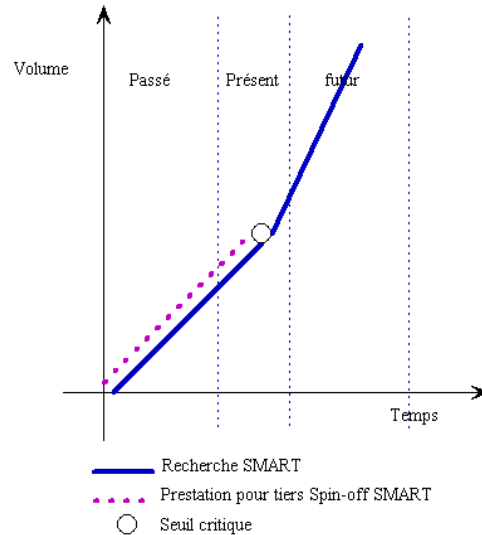
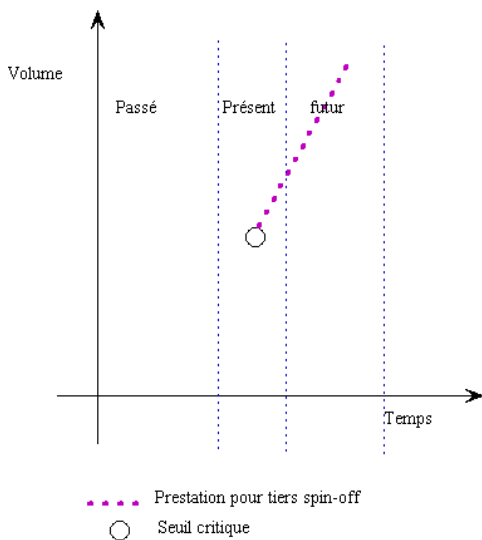


Soit répondre à la demande entrepreneuriale en diminuant son volume de recherche (scénario 2).

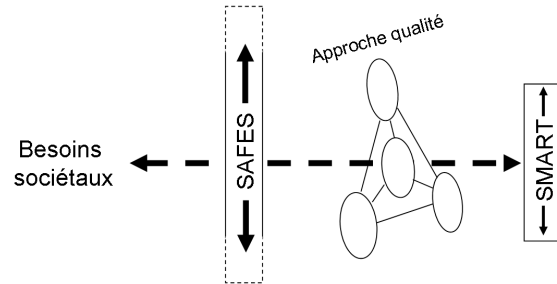


Nous sommes convaincus qu'en créant la spin off ASSESS GROUP, nous donnerons une chance aux prestations pour tiers de s'épanouir au sein de celle-ci tout en nous assurant que le SMART pourra s'investir avec plus de forces et de moyens dans la sphère scientifique.

Par cet entremise, le SMART assume pleinement sa responsabilité sociale d'ouverture et de création d'emplois dans un contexte socio-economico-démographique liégeois dans lequel le redéploiement économiquement est un leitmotiv.



Ce pari ne sera toutefois gagné que si les équipes, actuellement sous la même bannière mais bientôt séparées de fait, gardent un haut niveau de collaboration et d'échange. Ainsi, le SMART pourra renforcer ses R&D à l'aide d'une analyse des besoins mieux cernée par la spin-off et cette dernière pourra valoriser des outils et des méthodes répondant mieux aux besoins. A cette condition, ensemble, la spin off ASSESS GROUP et le SMART seront armés pour rencontrer les besoins sociétaux de recherche et de services aux entreprises.



A cette fin le SMART oeuvrera à assurer la qualité scientifique d'outils qui solutionneront des problèmes sociétaux émergents relayés, nous l'espérons, par l'ancrage de la spin-off des réalités de terrain.

Au delà recherche de la recherche appliquée, notre projet de spin-off répond aussi à des besoins sociaux en terme de création d'emplois par des chercheurs qui en ont la possibilité et donc montre une nouvelle facette du métier de la recherche qui intègre aussi de plus en plus les aspects valorisation de résultats

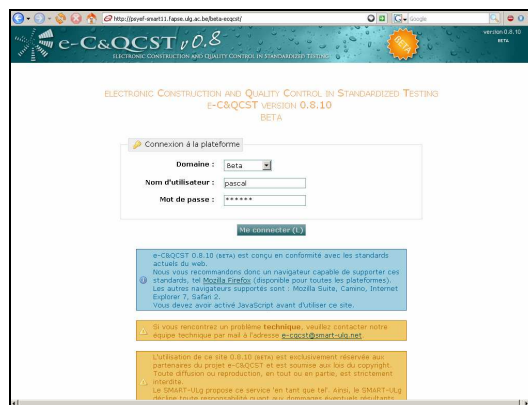
## **VI. OUTPUT DU PROJET : LA PLATE-FORME EXAMS**

Afin de vous faire découvrir la plate-forme « *examination Assessment Management System (exAMS)* » issue de ce projet R&D « *electronic Construction and Quality Control in Standardized Testing (e-C&QCST)* », vous trouverez dans les pages suivantes une copie du Quick Start Guide qui a été réalisé.

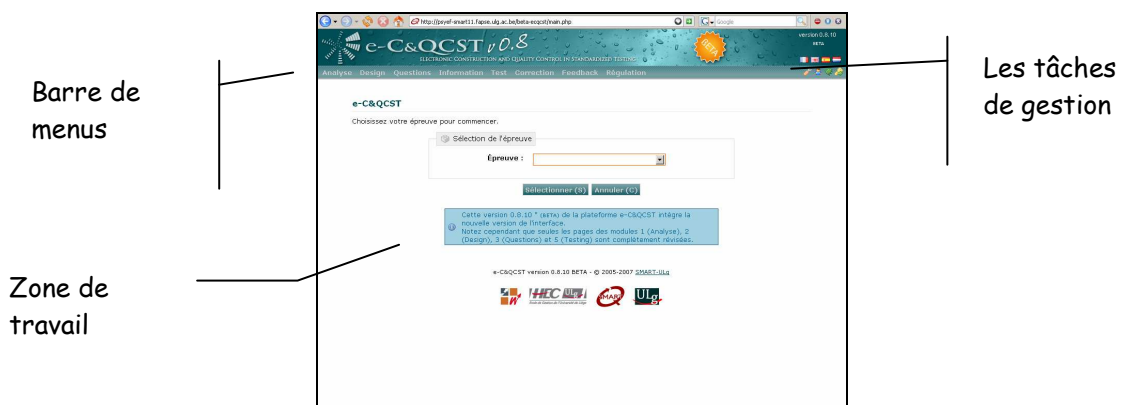
## Quick Start Guide de la plate-forme exAMS

La plate-forme exAMS (*examination Assessment Management System*) a pour objectif de mettre en place une procédure permettant aux évaluateurs de réaliser des évaluations standardisées de qualité. La plateforme est constituée d'un moteur informatique sur lequel est greffé une interface pour les concepteurs des évaluations et une interface pour les évalués. D'autres interfaces, comme l'interface opérateur, qui permettra à un opérateur d'aider l'enseignant dans la phase de correction, sont en cours de développement. L'interface des évaluateurs est actuellement constituée de 8 modules. La plate-forme comprend également un module d'administration. Chaque module de l'interface évaluateur est composé de sous-étapes. Pour réaliser un test standardisé de qualité, il est nécessaire de passer par chacun des modules et chacune des sous-étapes qui les composent. Dans les pages ci-dessous, vous allez trouver une description générale de la plateforme, puis l'essentiel des menus de l'interface utilisateurs, et ce, pour les 7 modules déjà réalisés. Enfin, seront illustrés le module « évalués » et l'interface « administration ».

### 1) L'organisation générale de la plate-forme



La première des choses à faire est de se connecter et d'encoder son identifiant et son mot de passe afin d'être reconnu par la plateforme. Une fois connecté à la plateforme, vous avez un aperçu du type d'interface qui vous guidera tout au long de votre travail. Celle-ci est constituée de barre de menus, d'icônes permettant d'accéder à des tâches de gestion et par un espace de travail.



## 1.1. Les icônes de tâches de gestion



L'administration de la plateforme (la partie administration est précisée au chapitre 3 de ce présent Quick Start Guide)

Vous permet de créer vos épreuves, vos groupes d'évalués ainsi que vos taxonomies et de compléter les données relatives à votre profil.

Les sous-étapes du module sont :

- La gestion des évalués
- La gestion des taxonomies
- La gestion des épreuves
- La gestion de son profil



Pour quitter le module administration, cliquez sur  et vous retournez directement à la sous-étape à laquelle vous étiez auparavant.



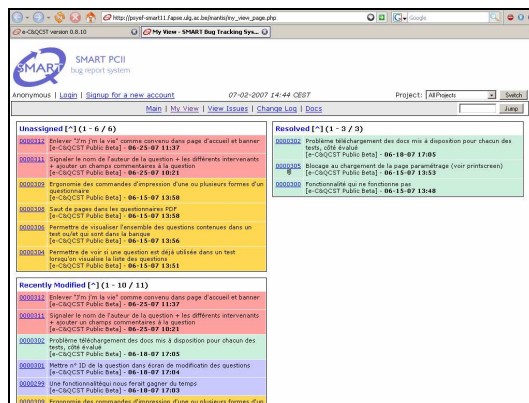
**Votre profil**

Vous permet d'accéder à une zone d'information qui vous renseigne sur votre nom, votre institution, votre fonction, votre département, votre type d'utilisateur et votre rôle. Vous pouvez également identifier les autres utilisateurs connectés à la plateforme.



**Le bug tracking**

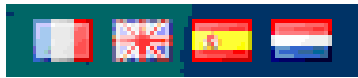
Vous permet de communiquer d'éventuels dysfonctionnements à l'équipe de développeurs.





Quitter la plateforme

Vous permet de quitter la plateforme.



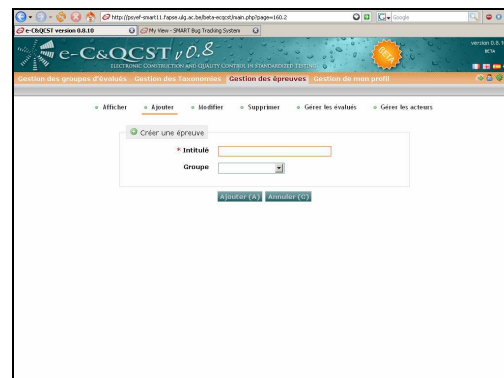
Multilinguisme

Vous permet de choisir votre langue de travail.

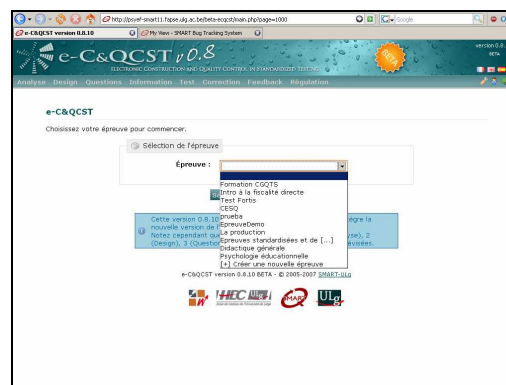
## 1.2. La barre de menus

La première étape à effectuer pour accéder à l'ensemble de la barre de menus est de créer ou choisir une épreuve, c'est-à-dire l'élément le plus englobant que nous pouvons faire correspondre à un cours, une formation, un référentiel de compétences ou à un métier qui sera évalué.

Pour créer une épreuve, vous devez vous rendre dans le module administration de la plateforme et choisir le menu « gestion des épreuves ». En cliquant sur ajouter, vous ouvrez un menu dans lequel vous pouvez encoder un nom d'épreuve et y rattacher un groupe d'évalués.



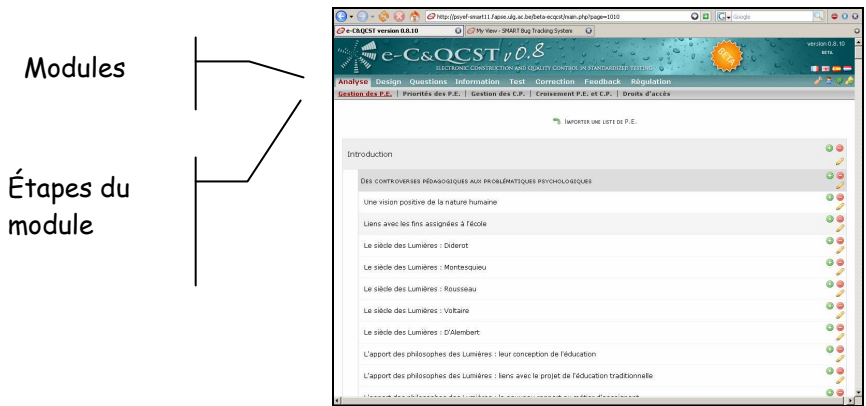
Pour sélectionner une épreuve créée précédemment (créée par vous-même ou pour laquelle on vous a donné des droits d'accès), choisissez-la dans le menu déroulant présenté au centre de la page initiale qui vous est présentée après identification.



## 2) Les modules et les sous-étapes

L'interface « évaluateur » est constituée de 8 modules. Chaque module est composé de sous-étapes.

Vous pouvez voir les huit modules dans le menu horizontal en haut de l'écran. Le module dans lequel vous travaillez apparaît en rouge. Lorsque vous êtes dans un module, les sous-modules attachés à celui-ci sont disponibles juste en dessous du menu. A nouveau, le sous-module affiché apparaît en rouge.



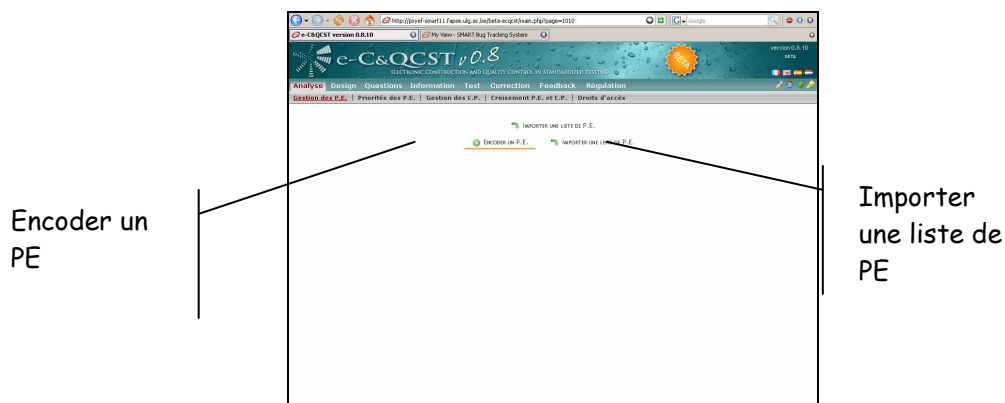
### 2.1. Le module Analyse

Ce module vous permettra de créer votre table de spécification et d'associer aux points à évaluer les catégories de performance que vous souhaitez exploiter.

#### Gestion des PE

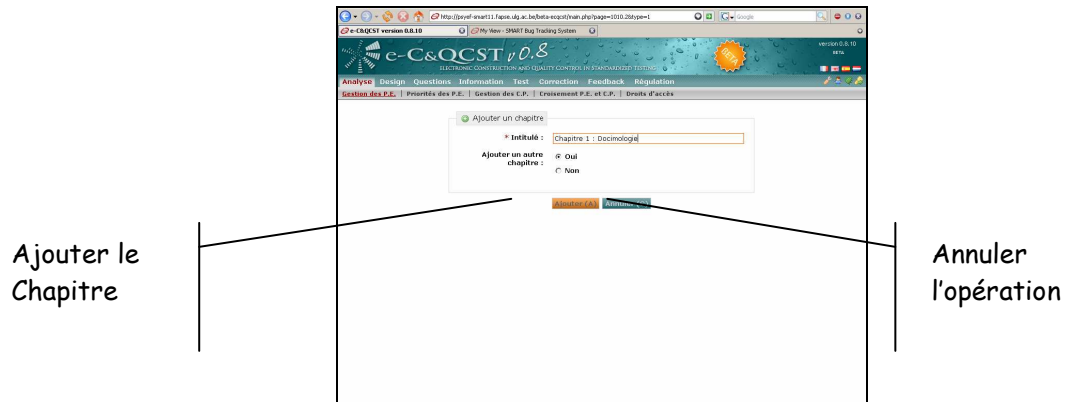
A cette étape, vous encodez la table de spécification de votre épreuve.

Au départ, vous avez le choix entre encoder un PE (premier scénario) ou importer une table de spécification (second scénario).



Si vous choisissez « encoder un PE », l'interface présentée vous permet d'encoder un chapitre. En effet, un Point à Evaluer (PE) est toujours rattaché à un sous-chapitre *père*, lui-même rattaché à un chapitre *père* :



- Intitulé : l'intitulé donné à votre chapitre apparaîtra dans la table de spécification.
- Ajouter un autre chapitre : "oui" sélectionné permet d'ajouter un chapitre supplémentaire, sinon vous arrivez à votre début de table de spécification.





Cliquez sur Ajouter (A) pour confirmer votre ajout, sur Annuler (C) dans le cas contraire.

Ensuite, vous pouvez encoder un(ou des) chapitre(s), sous-chapitre(s) et enfin, PE à partir des icônes suivantes :



-  Vous permet d'encoder un élément fils par rapport à l'élément sélectionné (par exemple, un sous-chapitre si aucun n'a encore été encodé dans un chapitre) dans votre table de spécification.
-  Vous permet d'encoder un élément de même niveau dans votre table de spécification. Vous pourrez choisir la position de cet élément grâce au menu déroulant et les boutons Avant ou Après.



- 
 Vous permet de supprimer un élément préalablement encodé dans la table de spécification.
- 
 Vous permet d'éditer et donc de modifier l'intitulé ou la position d'un élément de la table de spécification.

## Priorités des PE

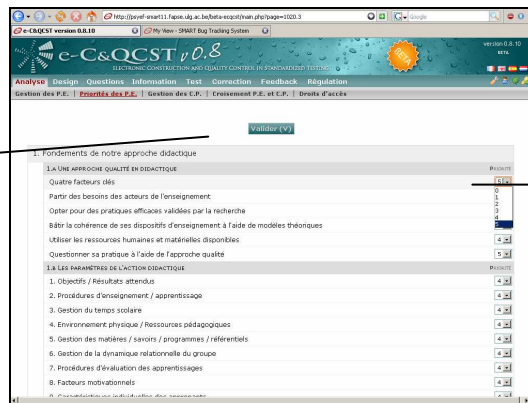
A cette étape, vous associez des priorités aux PE encodés à l'étape précédente.



Cliquez sur Modifier.

Utilisez le menu déroulant mis à votre disposition en face de chaque PE et sélectionnez la priorité souhaitée.

Modifier



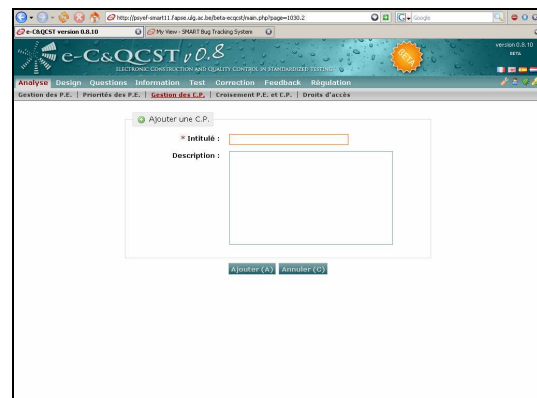
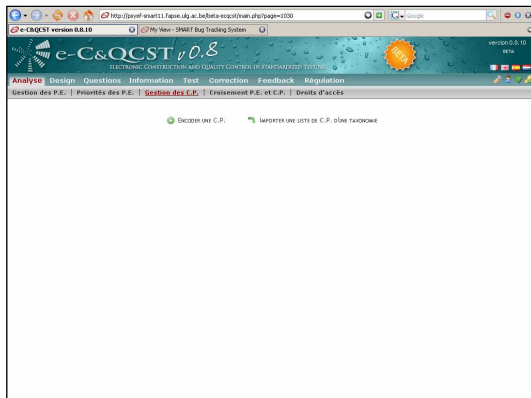
Menu de sélection de priorités

Une fois toutes les priorités encodées, cliquez sur Modifier (E) afin de valider votre choix de priorités.

### Gestion des CP

Cette étape permet de créer ou d'importer les catégories de performance que vous souhaitez exploiter dans votre épreuve.

Au départ, vous avez le choix entre importer une CP d'une taxonomie préexistante ou d'encoder directement les Catégories de performances.




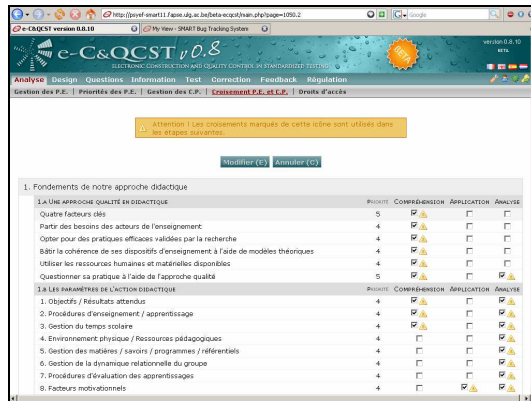
Si vous encodez directement la catégorie de performance, vous devez donner un nom et une description à la CP

### Croisement PE et CP

Cette étape a pour but de croiser les points à évaluer avec les catégories de performance que vous pensez légitime d'exploiter pour ce PE.

Un tableau présentant les différents croisements possibles est exposé lorsque vous cliquez sur modifier. Si certains croisements déjà activés sont utilisés dans la suite du cycle et ne peuvent être

supprimés sans dommage, ils sont spécifiés par l'icône .

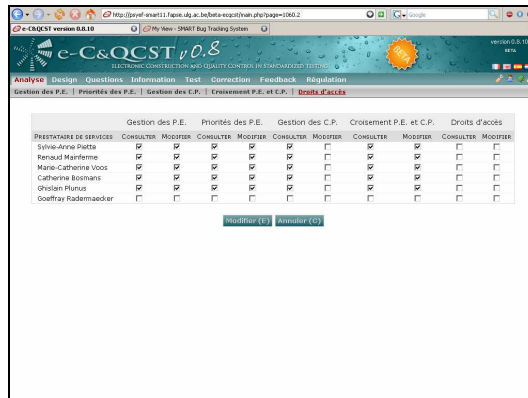


Vous ajoutez des croisements simplement en cochant les cases à cocher.

Cliquez sur Modifier (E) pour valider votre sélection, sur Annuler (C) dans le cas contraire.

### Droits d'accès

Cette section permet d'accorder des droits (de lecture (consultation) et/ou écriture (modification)) à des acteurs choisis pour chaque étape présente dans le module.



## 2.2. Le module Design

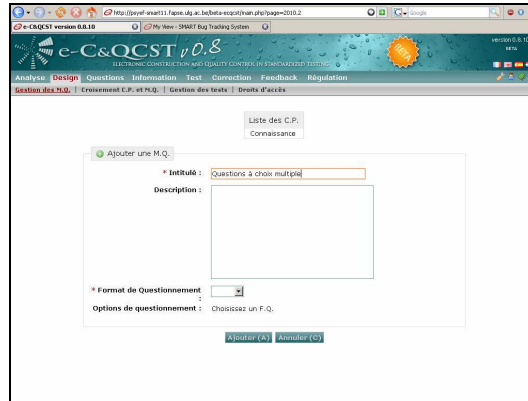
Ce module vous permettra de mettre en forme votre épreuve, en choisissant les modalités de questionnement que vous voulez employer, en créant des tests et en sélectionnant pour chacun d'eux les trinômes constitués des [Points à évaluer]\*[Catégories de performance]\*[Modalités de questionnement].

### Gestion des MQ

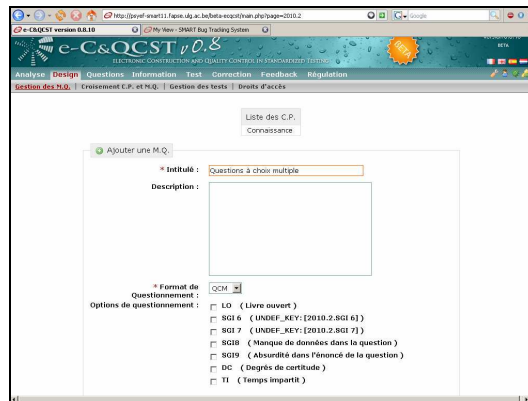
Cette étape permet de gérer les modalités de questionnement que vous souhaitez utiliser dans vos tests. Lorsque vous cliquez sur « encodez des MQ » vous devez encoder les champs suivants :

- Intitulé : il apparaît dans le tableau présentant les différentes MQ existantes ;

- Description : vous permet de donner une description complète de votre MQ ;
- Format des Questionnements : vous permet de sélectionner le format de questionnement pour votre MQ. Vous avez le choix entre VF, VFG, QCM, QCL et autres.






Lorsque vous choisissez un « Format de questionnement », l'interface vous propose de choisir des options de questionnement qui peuvent y être associées



Cliquez sur Ajouter (A) pour confirmer votre ajout, sur Annuler (C) dans le cas contraire.

Lorsqu'au moins une Modalité de questionnement a été encodée, vous pouvez gérer la liste des MQ.

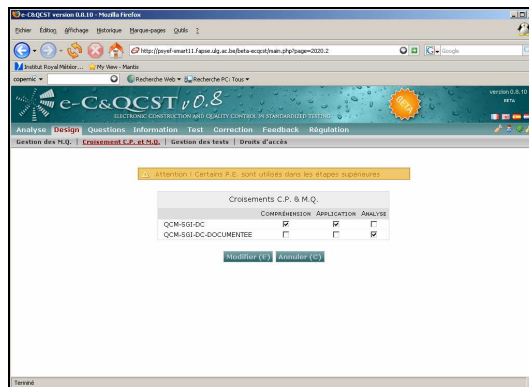


-  Vous permet d'ajouter une MQ.
-  Vous permet de supprimer une MQ déjà encodée.
-  Vous permet d'éditer puis de modifier l'intitulé, la description, le format de questionnement ou l'option de questionnement d'une MQ existante.

## Croisement CP et MQ

Cette étape permet de convenir des croisements que vous souhaitez obtenir entre les catégories de performance et les modalités de questionnement.

Quand vous cliquez sur modifier, un tableau présentant les différents croisements possibles est exposé.

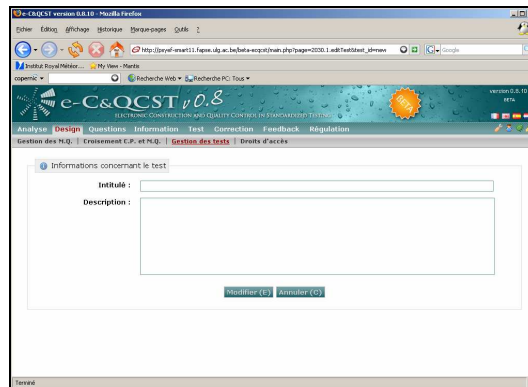


Pour sélectionner un croisement, cochez la case à cocher correspondante.

Cliquez ensuite sur Modifier (E) pour valider votre sélection, ou sur Annuler (C) pour ne pas enregistrer vos modifications.

## Gestion des tests

Cette étape vous permet de créer, modifier ou supprimer le ou les test(s) associé(s) à votre épreuve. Cliquez sur Créez un nouveau test. Vous obtenez l'écran suivant :

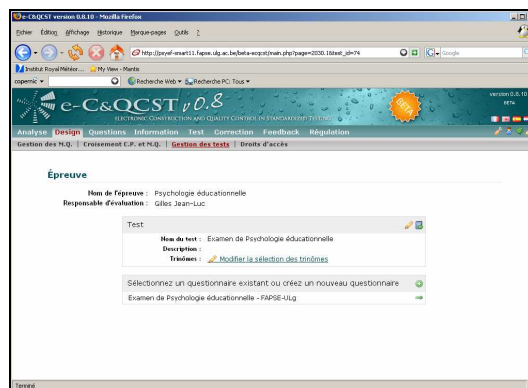


Encodez les champs suivants :






- Intitulé : vous permet de nommer le test.
- Description : vous permet de donner une description complète de votre test.

Cliquez sur Ajouter (A) pour confirmer votre ajout, sur Annuler (C) dans le cas contraire.

L'écran suivant est alors affiché :



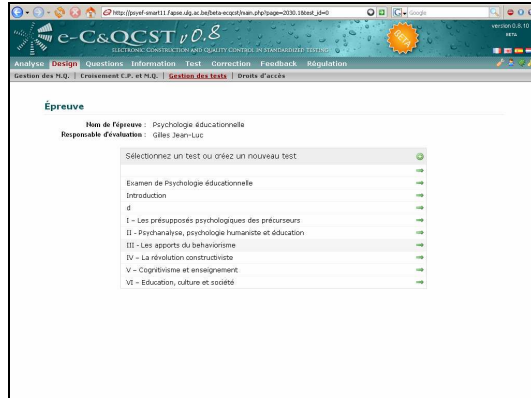
A ce stade vous pouvez :


-  éditer et éventuellement modifier les informations concernant le test ;
-  quitter ce menu et revenir à la liste des tests ;
-  [Modifier la sélection des trinômes](#) modifier la sélection des trinômes associés à ce test ;
-  sélectionner un questionnaire existant ou créer un nouveau questionnaire pour ce test ;
-  si une passation de ce test a déjà eu lieu, consulter son nom, son type et sa description.

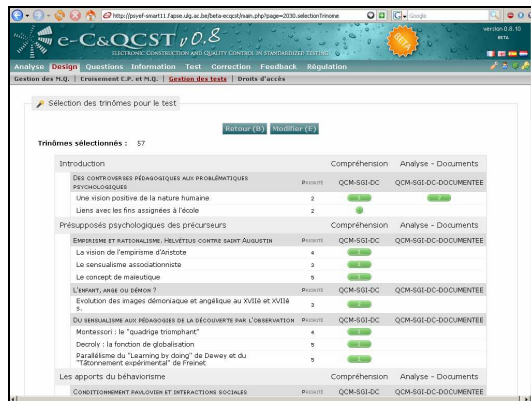
## Sélectionner/modifier la liste des trinômes utilisés pour le test

Cette étape vous permet de sélectionner/modifier les trinômes (c'est-à-dire la combinaison PE-CP-MQ) que vous souhaitez associer à votre test.

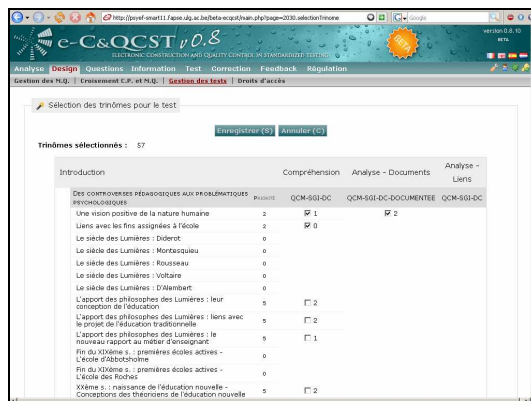
Sélectionnez dans la liste des tests le test auquel vous souhaitez associer des trinômes.



Le lien  [Modifier la sélection des trinômes](#) vous permet d'exposer votre table de spécification.




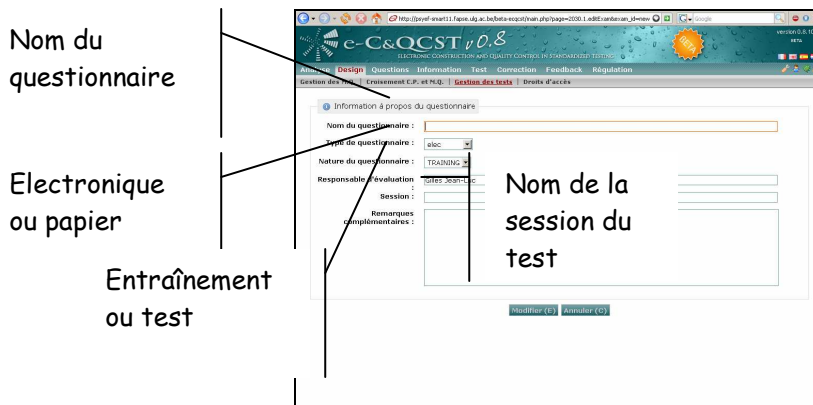
Lorsque vous cliquez sur modifier, la table de spécification apparaît avec des cases sélectionnables situées aux combinaisons PE-CP-MQ potentielles, suite aux choix réalisés aux étapes précédentes.



Cliquez sur Modifier (E) pour enregistrer vos modifications sinon sur Annuler (C).

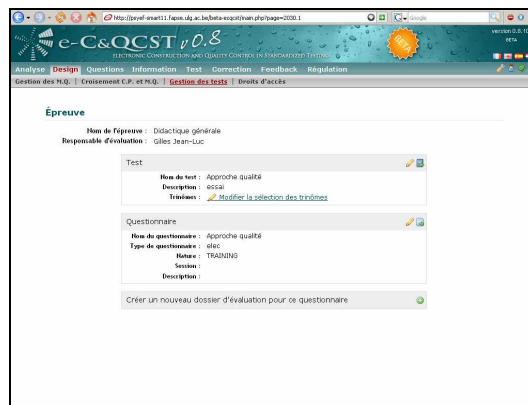
**Sélectionnez un questionnaire existant ou créez un nouveau questionnaire**



Lorsque vous choisissez l'option  **Sélectionnez un questionnaire existant ou créez un nouveau questionnaire pour ce test, l'écran suivant apparaît :**



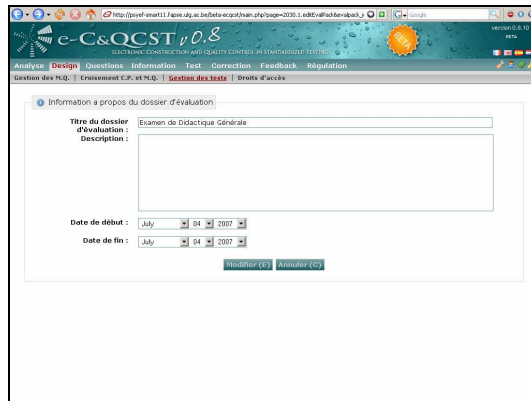
*Vous devez remplir ou modifier les champs de cette page pour créer ou modifier un questionnaire.*

*Lorsque vous cliquez sur modifier, l'écran suivant apparaît :*

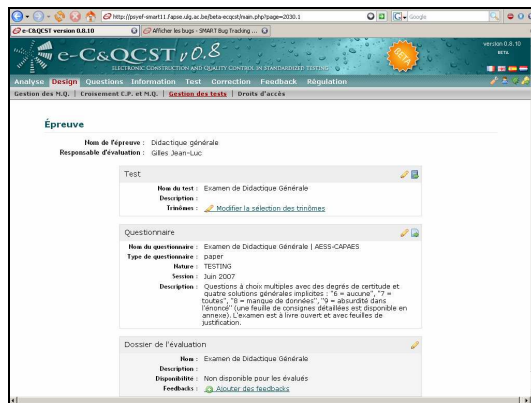


A partir de cet écran vous pouvez  **éditer chacune des options pour les modifier ou utiliser le bouton**  **pour quitter le menu et revenir en arrière.**

*Vous pouvez également sélectionner le menu « Créer un nouveau dossier d'évaluation pour ce questionnaire ». Vous obtenez alors l'écran suivant :*

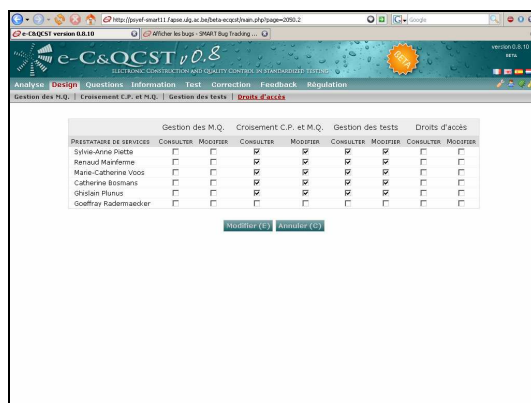


*A partir de celui-ci, vous pouvez donner un nom à votre dossier d'évaluation, le décrire et attribuer une période de mise à disposition pour les étudiants. Cliquez alors sur Modifier (E) pour enregistrer vos modifications sinon sur Annuler (C). Si vous cliquez sur modifier, les caractéristiques du dossier d'évaluation apparaissent. A partir de ce moment, vous pourrez, quand le test aura eu lieu, mettre à disposition les feedbacks étudiants en cliquant sur le lien « ajouter des feedbacks étudiants ».*



## Droit d'accès

Cette section permet d'accorder des droits (de lecture et/ou d'écriture) à des acteurs choisis pour chaque étape présente dans le module.



*Lorsque des acteurs ou groupes sont liés à l'épreuve, un tableau apparaît permettant de sélectionner les droits que vous souhaitez accorder à un ou plusieurs acteurs et/ou groupes.*



*Cliquez sur Modifier (E) pour valider votre sélection, sur Annuler (C) dans le cas contraire.*

### 2.3. Le module Questions

Au terme de ce module, une banque de questions relative à une épreuve sera validée et par conséquent, disponible pour être intégrée au test.

Le premier écran attaché à ce module vous permet de visualiser toute une série d'informations sur les tests et les questionnaires existants.



Les boutons  et  vous permettent d'approfondir ou inversement de résumer l'information disponible pour le test ou le questionnaire.

### Gestion des questions

Cette étape permet de gérer votre banque de questions en ajoutant, supprimant, modifiant des questions qui sont chacune rattachées à un trinôme [Point à Evaluer]\*[Catégorie de Performance]\*[Modalité de Questionnement]. (Les trinômes ont été constitués dans les modules 1 et 2). L'écran qui permet ce travail sur les questions est le suivant :

**Sélection du test**

**Rechercher une question par son ID**

**Liste des questions**

**Table de spécification de l'épreuve**

Compétence	Compréhension	Application	Analyse
1.1. LA LIGNE APPROCHE QUALITÉ EN DIDACTIQUE	Présente	QCM-SG-DG	QCM-SG-DC
Quelle fonction elle	5		
Pour des besoins des acteurs de l'enseignement	4		
Opter pour des pratiques efficaces validées par la recherche	4		
Bâter la cohérence de ses dispositifs d'enseignement à l'aide de modèles théoriques	4		
Utiliser les ressources humaines et matérielles disponibles	4		
Questionner sa pratique à l'aide de l'approche qualité	5		
1.1.2. LES ENVAHISSEMENTS DE L'ACTEUR DIDACTIQUE	Présente	QCM-SG-DC	QCM-SG-DC
1.1.2.1. Dispositif / Itinéraire alternatif	4		
2. Procédures d'enseignement / apprentissage	4		

Vous avez à ce stade 2 options.

[La première] est de cliquer sur le lien « liste de questions » ce qui vous fournira cette liste et vous permettra, en cliquant sur l'icône de visualiser chacune d'entre-elles. Vous pouvez également avoir accès à la visualisation des questions en tapant son numéro d'identifiant dans le champ « rechercher une question par son ID »

Text :

Liste des questions par trinôme pour le test :

Identifiant	Titre	Priorité	Statut
P.E. : DÉFINITION ET AVANTAGES DE LA PLANIFICATION		Priorité : 1	
C.P. : Compréhension			
M.Q. : QCM-SG-DC			
772	TMR - Modèle systématique de planification	3	
P.E. : PHASE RÉACTIVE / IMPORTANCE DES ROUTINES D'ACTIVITÉ, D'INTERVENTION, DE GESTION ET D'ÉVALUATION		Priorité : 3	
C.P. : Compréhension			
M.Q. : QCM-SG-DC			
729	Le rôle des élèves au sein des phases de l'action didactique	2	
752	TMR - Procédures automatisées d'intervention (routines)	2	
794	TMR - Planification des actions didactiques	2	
P.E. : PHASE RÉACTIVE / PLANIFICATION DE LA GESTION DE LA CLASSE : RÈGLES ET PROCÉDURES, REPRÉSENTATIONS ET ATTENTES		Priorité : 3	
C.P. : Compréhension			
M.Q. : QCM-SG-DC			
797	Les représentations et les attentes de l'enseignant	1	
817	Les représentations et les attentes de l'enseignant - version RC = 6	4	
P.E. : LES QUATRE ÉTAPES DE LA PLANIFICATION SELON TYLER		Priorité : 5	
C.P. : Compréhension			
M.Q. : QCM-SG-DC			
798	Modèle linéaire de Tyler	7	

Afficher la question

TMR - MIROIR SYSTÉMIQUE DE PLANIFICATION

Cette question a été validée

La citation de John W. Powell "La seule véritable erreur est celle dont on ne retire aucun enseignement" signifie, dans le modèle systématique de planification que :

1. La spécification des objectifs n'est plus aussi nécessaire que dans le modèle linéaire de planification.
2. Les erreurs commises lors de la planification ne peuvent être analysées que lors de la phase interactive.
3. La véritable erreur est celle qui consiste à ne pas tenir compte de ce qui n'a pas marché pour rectifier la prochaine action.
6. Solution Générale Implicite 6
7. Solution Générale Implicite 7
8. Solution Générale Implicite 8
9. Solution Générale Implicite 9

[La seconde] est de choisir le test sur lequel vous souhaitez travailler. Dans ce cas, la table de spécification inhérente de ce test vous est proposée. Elle peut vous être fournie en vue complète ou en vue simplifiée. L'option « table de spécification en vue complète » vous offre l'ensemble de la table de l'épreuve. Les trinômes sélectionnés pour le test apparaissent en vert (avec un chiffre éventuel correspondant au nombre de questions disponibles pour ce trinôme). Ceux qui n'ont pas été sélectionnés pour ce test apparaissent en rouge. L'option « table de spécification en vue simplifiée » vous propose la même table de spécification, mais les [points à évaluer] non sélectionnés ne sont pas affichés.

Modèle	Présent	QCM-S01-DC	QCM-S01-DC	QCM-S01-DC	QCM-S01-DC-DOCUMENTEE
1.0 LE TRIANGLE DIDACTIQUE : ENSEIGNANT-SAVOIR-APPRENANT	Présent	QCM-S01-DC	QCM-S01-DC	QCM-S01-DC	QCM-S01-DC-DOCUMENTEE
1.1 Les trois pôles (Hamelin)	5	●	●	●	●
1.2 L'opération didactique d'un enseignement trop transmisif (De Ketele, Delory)	4	●	●	●	●
1.3 L'opération didactique d'un enseignement trop "non-didactif" (De Ketele, Delory)	4	●	●	●	●
1.4 L'opération didactique d'un enseignement trop focalisé sur les objectifs (De Ketele, Delory)	4	●	●	●	●
1.5 Vers un triangle didactique plus équilibré	4	●	●	●	●
1.6 Influence des modèles d'enseignement transmissifs (Dewey, Cournot, Rogers, Oury, Legrand, Bloom, Wysocki, Piaget, Bruner, ...)	3	●	●	●	●
1.0 LE MODÈLE MOTIVATIONNEL DE VIAU	Présent	QCM-S01-DC	QCM-S01-DC	QCM-S01-DC	QCM-S01-DC-DOCUMENTEE
1.1 Causes de démotivation chez les étudiants universitaires	3	●	●	●	●
1.2 Influence des activités pédagogiques sur la motivation	3	●	●	●	●
1.3 Mécanismes conditionnels pédagogiques pour booster et maintenir la motivation des élèves	3	●	●	●	●
1.4 Lien entre le modèle de Viau et les postures des actions didactiques	3	●	●	●	●
2. Planification didactique : modèles théoriques et procédures		Compréhension	Application	Analyse	
2.1 PLANIFICATION DIDACTIQUE ET RECHERCHE SUR LES PRATIQUES	Présent	QCM-S01-DC	QCM-S01-DC	QCM-S01-DC-DOCUMENTEE	
2.1.1 Définition et avantages de la planification	1	●	●	●	●
2.1.2 Phase préactive / Importance des routines d'activité, d'intention, de gestion et d'évaluation	3	●	●	●	●
2.1.3 Phase préactive / Planification de la gestion de la classe : règles et procédures, représentations et attitudes	3	●	●	●	●
2.1.4 Phase préactive / Gestion du transfert des apprentissages : liens, éléments contextuels et similaires des situations	3	●	●	●	●
2.1.5 SÉRIERER LES OBJECTIFS DANS UNE APPROCHE LINÉAIRE DE					

Question	Présent	Compréhension	Application	Analyse
1. Fondements de notre approche didactique	Présent	QCM-S01-DC	QCM-S01-DC	QCM-S01-DC-DOCUMENTEE
1.A UNE APPROCHE QUALITE DE DIDACTIQUE	Présent	QCM-S01-DC	QCM-S01-DC	QCM-S01-DC-DOCUMENTEE
1.A LES PARAMETRES DE L'ACTION DIDACTIQUE	Présent	QCM-S01-DC	QCM-S01-DC	QCM-S01-DC-DOCUMENTEE
1.C LE TRIANGLE DIDACTIQUE : ENSEIGNANT-SAVOIR-APPRENANT	Présent	QCM-S01-DC	QCM-S01-DC	QCM-S01-DC-DOCUMENTEE
1.C LE TRIANGLE DIDACTIQUE : ENSEIGNANT-SAVOIR-APPRENANT	Présent	QCM-S01-DC	QCM-S01-DC	QCM-S01-DC-DOCUMENTEE
1.C LE TRIANGLE DIDACTIQUE : ENSEIGNANT-SAVOIR-APPRENANT	Présent	QCM-S01-DC	QCM-S01-DC	QCM-S01-DC-DOCUMENTEE
2. Planification didactique : modèles théoriques et procédures		Compréhension	Application	Analyse
2.1 PLANIFICATION DIDACTIQUE ET RECHERCHE SUR LES PRATIQUES	Présent	QCM-S01-DC	QCM-S01-DC	QCM-S01-DC-DOCUMENTEE
2.1.1 Définition et avantages de la planification	1	●	●	●
2.1.2 Phase préactive / Importance des routines d'activité, d'intention, de gestion et d'évaluation	3	●	●	●

L'icône vous informe sur légende relative à la table de spécification.

- Ce trinôme P.E - C.P. - M.Q. est sélectionné
- Ce trinôme P.E - C.P. - M.Q. est sélectionnable
- Ce trinôme P.E - C.P. - M.Q. n'est pas sélectionné
- Ce P.E. ne possède aucun croisement et n'est pas sélectionnable.
- Ce P.E. est sélectionné

Une fois que vous avez cliqué sur l'icône du trinôme sélectionné ou sélectionnable ' derrière lequel vous souhaitez ajouter une question, vous obtenez l'écran suivant :

**Caractéristiques du trinôme**





**Liste des questions attachées à ce trinôme**

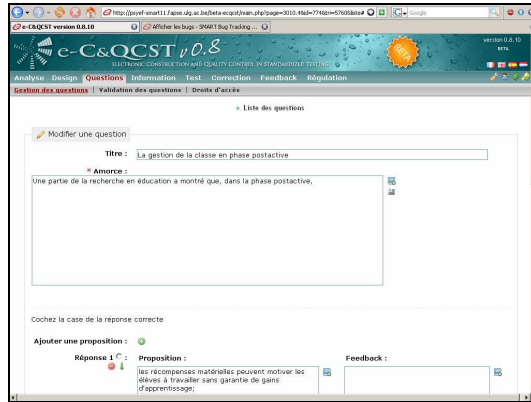
**Ajouter une question**




L'icône vous permet de visualiser la question.

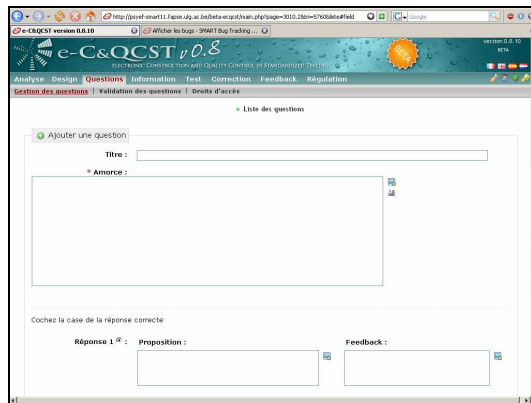
L'icône vous permet de cloner une question.

L'icône vous permet de supprimer une question.

L'icône  vous permet d'éditer la question pour la modifier. Vous pourrez alors changer son titre, son amorce, ses distracteurs et ses feedbacks. L'icône  vous permet de supprimer un distracteur et le feedback qui lui est associé. Les icônes   vous permettent de modifier la place des distracteurs.






L'icône  vous permet d'ajouter une question. L'écran suivant vous permet alors d'encoder les attributs (amorce, distracteurs, feedbacks, réponse correcte). L'icône  vous permet d'insérer du multimédia dans l'amorce, dans les propositions de réponse et dans les feedbacks. L'icône  vous permet de travailler la mise en page de l'amorce.



Enfin, lorsque vous cliquez sur le bouton « Liste des questions » vous obtenez la liste des questions relatives à l'épreuve ou au test sélectionné.

## Validation des questions

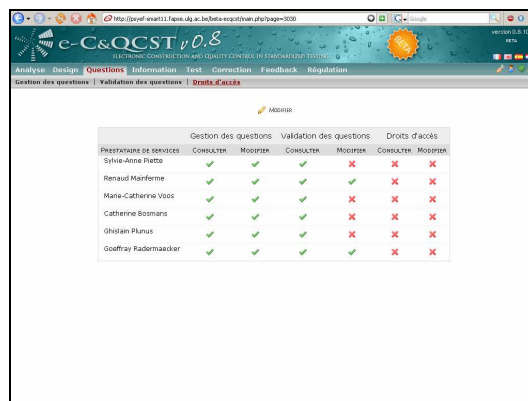
Cette étape permet de valider les questions encodées afin qu'elles puissent faire partie de la passation d'un test.

Les premières phases sont identiques à celles de la gestion des questions. Vous devez d'abord sélectionner un trinôme sélectionné  ou sélectionnable . La liste des questions attachées à ce trinôme apparaît. Vous pouvez visualiser la question à l'aide de l'icône . Si vous souhaitez la valider, vous devez cocher la case à cocher correspondante.



## Droits d'accès

Cette section permet d'accorder des droits (de lecture (consultation) et/ou écriture (modification)) à des acteurs choisis pour chaque étape présente dans le module.



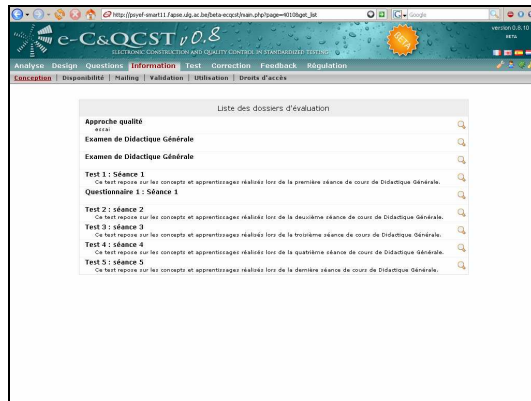
Cliquez sur Modifier (E) pour valider votre sélection, sur Annuler (C) dans le cas contraire.


## 2.4. Le Module Information

Ce module permet la mise en place de "modules d'évaluation" composés d'un examen (la concrétisation d'un test) mais également d'un ensemble d'éléments associés à celui-ci tels que des documents informatifs, des tests d'entraînement, des informations internes à la plateforme (par exemple, la table de spécification), des liens vers de la documentation externe,...

### Conception

La page d'accueil de cette étape présente la liste de toutes les évaluations de l'épreuve (en se référant aux tests créés à l'étape "Gestion des tests" du module "Design").



Pour configurer un dossier d'évaluation que vous avez créé à l'étape 2 design et gérer l'ensemble des éléments associés à celui-ci, vous devez cliquer sur l'icône  du test concerné. Vous obtenez l'écran suivant :

Ajouter des éléments dans le dossier d'évaluation

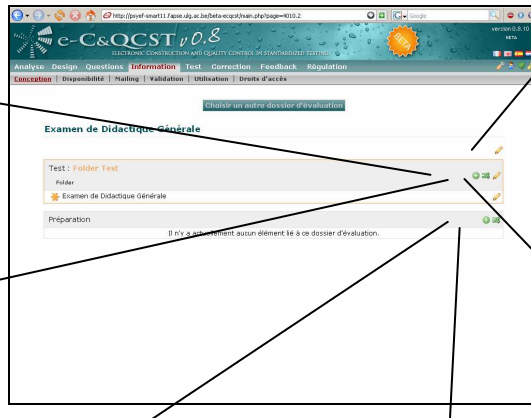
Modifier l'ordre de présentation des éléments du dossier d'évaluation

Ajouter des éléments à la préparation de l'évaluation

Modifier l'ordre de présentation des éléments de la préparation

Modifier le nom et les commentaires du dossier d'évaluation

Modifier le nom et les commentaires des éléments du dossier d'évaluation



Deux sous-ensembles relatifs à l'évaluation se distinguent. Le premier, intitulé "Test", se rapporte à la passation du test ; le second, nommé "Préparation", à la préparation de cette passation.

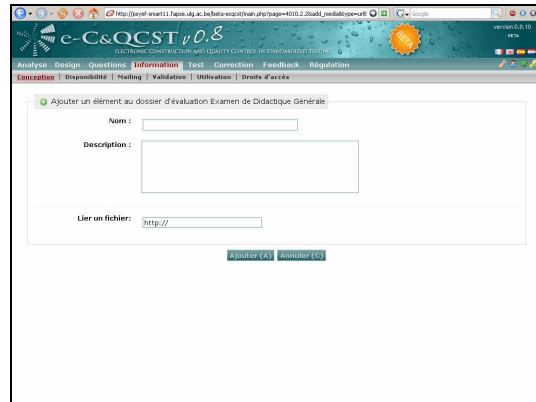
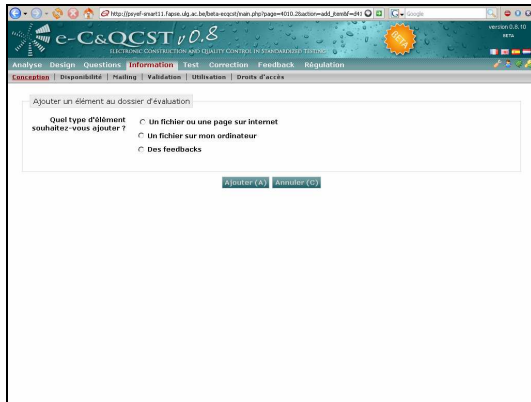
Le premier sous-ensemble "Test" représente le dossier comprenant le test ainsi que les fichiers qui y sont attachés.

Dans ce sous-ensemble, vous gérez le dossier (modification de son nom et de sa description), le test lui-même (modification de son nom, de sa description, de son type) ainsi que les éléments associés à cette passation (ajouter, modifier, supprimer un élément) et ce, à partir des icônes suivantes :



Vous permet d'ajouter des fichiers relatifs à votre passation. A ce stade, deux types de fichiers peuvent être attachés : un fichier ou une page Internet ainsi qu'un fichier provenant de votre ordinateur. Ces fichiers seront disponibles durant la passation de votre test. En cliquant sur cette

icône, l'écran suivant apparaît. En choisissant une des options, vous pouvez encoder les informations y afférentes.



Vous permet de gérer l'ordre de présentation des éléments attachés à votre épreuve.



Vous permet d'éditer, de modifier les éléments associés (nom, description,...) d'abord au dossier Test, ensuite au test et enfin, aux fichiers attachés.

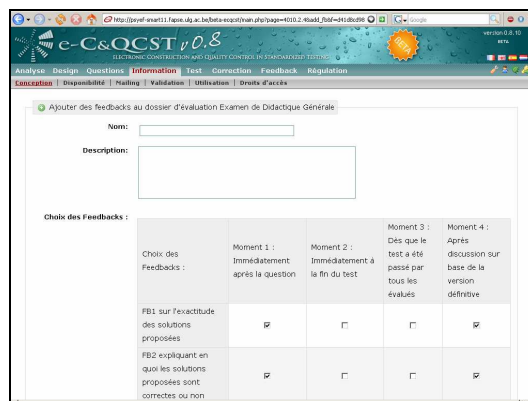


Vous permet de supprimer un élément que vous auriez préalablement attaché (le test et les éventuels feedbacks associés ne peuvent être supprimés).

Deuxième sous-cadre "Préparation" : il est composé de l'ensemble des éléments que vous souhaitez mettre à la disposition des évalués, avant la passation du test, en guise de préparation.



Vous permet d'associer des fichiers. Vous pouvez ajouter un fichier ou une page Internet, un fichier provenant de votre ordinateur, un test d'entraînement ou encore un folder. Ce dernier élément vous permet, lorsque vous souhaitez associer un grand nombre d'éléments à votre préparation de test, de créer un dossier dans lequel vous pourrez y intégrer différents fichiers (de la même manière que vous regroupez vos fichiers dans des répertoires). C'est également ici que vous pouvez associer des Feedbacks au test. Quand à la question « Quel type d'éléments souhaitez-vous ajouter », vous répondez « des feedbacks », le tableau à double entrée suivant s'affiche. Dans celui-ci, vous pouvez choisir le type de feedbacks souhaité et le moment de présentation du Feedbacks



Choix des Feedbacks :	Moment 1 : Immédiatement après la question	Moment 2 : Immédiatement à la fin du test	Moment 3 : Dès que le test a été passé par tous les évalués	Moment 4 : Après discussion sur base de la version définitive
FB1 sur l'exactitude des solutions proposées	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
FB2 expliquant en quoi les solutions proposées sont correctes ou non	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>



Vous permet de gérer l'ordre de présentation des fichiers attachés à votre épreuve.



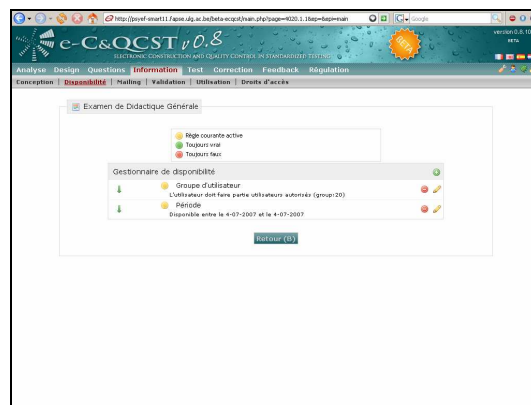
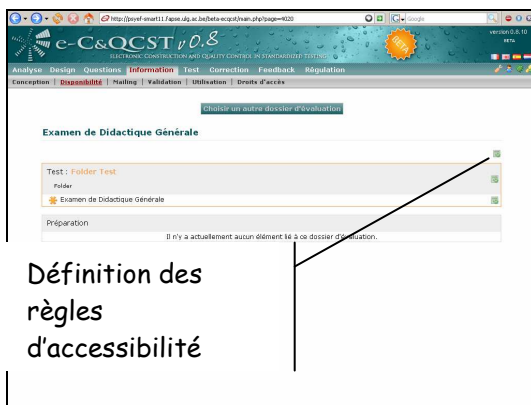
Vous permet de modifier le nom, la description, l'adresse Internet ou l'emplacement d'un fichier ou le contenu d'un folder.




Vous permet de supprimer l'élément attaché relatif à l'icône.

## Disponibilité

Des règles de mise à disposition des dossiers d'évaluation peuvent être associées au niveau de l'évaluation (par exemple, qui peut y avoir accès, pendant quelle période) dans sa globalité, au niveau du dossier Test (test en lui-même et des fichiers qui y sont attachés), ainsi qu'au niveau du dossier de la préparation à la passation. Attention, une règle définie au niveau du dossier d'évaluation s'impose automatiquement au niveau test et au niveau passation. En d'autres mots, si l'on décide que tout le dossier sera disponible pour tels étudiants à tels moments, cela signifie que tous les éléments du dossier seront accessibles à ce moment préalablement défini.



Deux règles de disponibilités sont actives par défaut. D'autres peuvent être ajoutées via l'icône . Les deux règles proposées par défaut sont [1] une règle relative au groupe d'utilisateurs, [2] une règle de période de mise à disposition.

Pour chaque règle, trois comportements sont proposés :




Règle courante active. La règle est active lorsque les conditions édictées par celle-ci sont remplies.






Toujours vrai. Quoiqu'il arrive, le logiciel interprète la condition comme remplie, même si dans les faits elle ne l'est pas.



Toujours faux. Quoiqu'il arrive, le logiciel interprète la condition comme non remplie, même si dans les faits elle l'est.

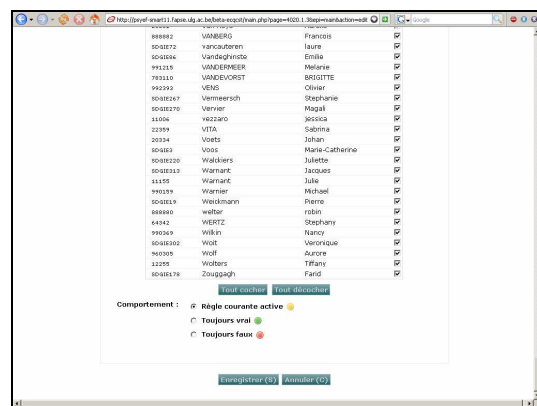
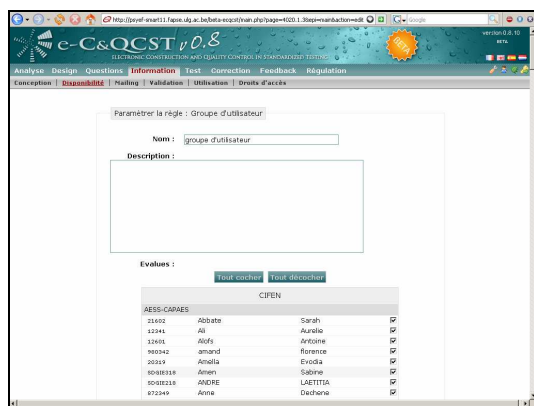
L'icône  signifie que la condition est considérée comme remplie. Soit la condition l'est réellement, soit l'utilisateur l'a notifiée comme étant toujours vraie. Si la condition n'était pas considérée comme

remplie, soit qu'elle ne l'est effectivement pas, soit que l'utilisateur l'a notifiée comme étant toujours fautive, l'icône serait .

L'icône  vous permet de supprimer la règle de la liste des règles attachées à votre dossier d'évaluation. L'icône  vous permet d'éditer la règle pour la configurer ou la modifier.

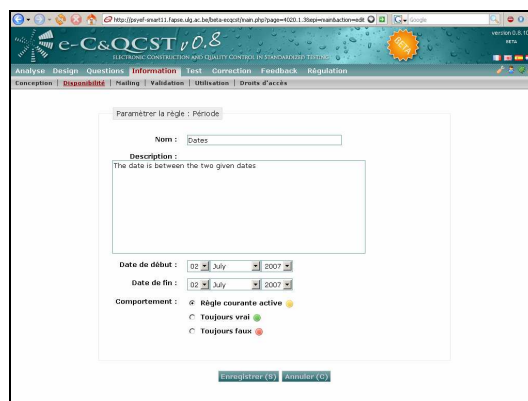
### Configurer ou modifier la règle « groupe d'utilisateurs »

Cette règle vous permet d'associer des évalués à votre évaluation dans sa globalité, à la passation de votre test ainsi à la préparation de celle-ci. La liste des évalués est celle qui a été introduite pour l'épreuve via l'interface administration. Vous pouvez donner un nom ainsi qu'une description à votre groupe d'utilisateurs. Ensuite, vous sélectionnez les évalués que vous souhaitez associer à votre test. Vous définissez ensuite le comportement attendu pour cette règle.



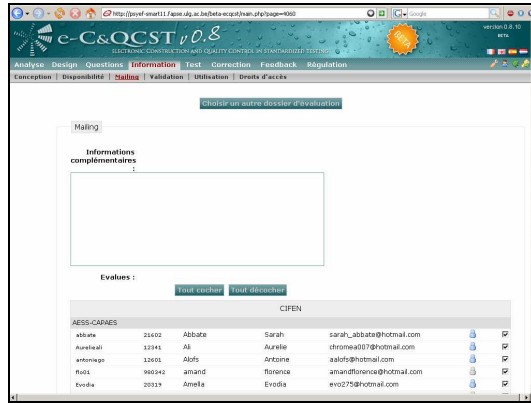
### Configurer ou modifier la règle « période » :

Cette règle vous permet de mettre à disposition les éléments soumis à la règle durant la période sélectionnée (date de début - date de fin).



## L'option mailing

Cette option permet d'envoyer un courriel au groupe d'évalués. Un champ est disponible pour le corps du message et la liste des évalués sélectionnés dans le dossier d'évaluation à l'étape précédente sont présélectionnés. Ils peuvent être désélectionnés au besoin.




## L'option validation

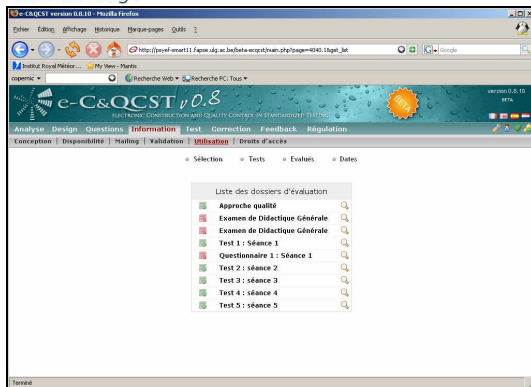
Lorsque cette option sera disponible, elle permettra de valider le dossier d'évaluation.

## L'option utilisation

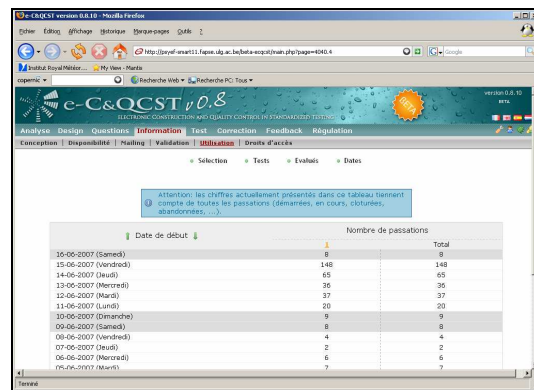
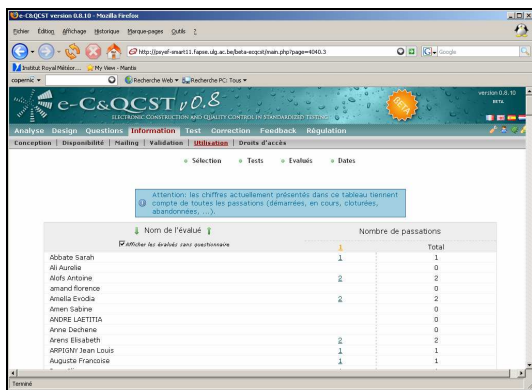
Cette option vous donne accès à une série de statistiques liées à la passation du test. Le premier écran vous donne le nombre de passations du test que vous avez sélectionné.



Pour avoir accès à la liste complète des tests de l'épreuve et ainsi pouvoir sélectionner les statistiques d'un autre test, vous devez cliquer sur le lien « sélection ». Cliquez alors sur l'icône  pour visualiser les statistiques du test en vis-à-vis.

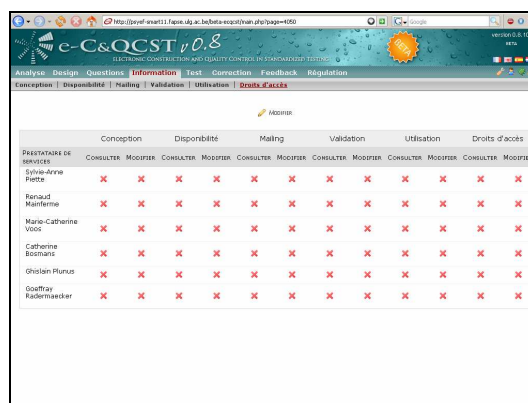


Le lien « Evalués » vous donne les statistiques sur le nombre de passation par évalué. Enfin, le lien « dates » vous donne les statistiques de passation du test en fonction du calendrier.



## Droits d'accès

Cette section permet d'accorder des droits (de lecture et/ou d'écriture) à des acteurs pour chacune des étapes présentes dans ce module.

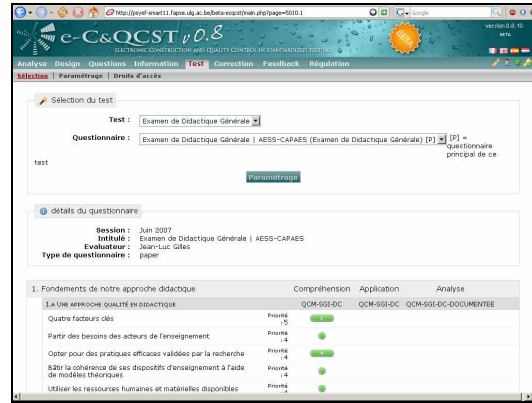
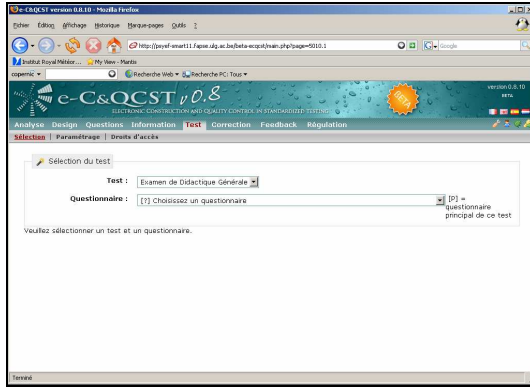


## 2.5 Le Module Testing

Ce module concerne le paramétrage de votre passation tant au niveau de l'examen lui-même (par exemple, le choix des questions ou l'ajout de consignes) qu'au niveau de sa correction.

## Sélection

A cette étape, vous sélectionnez le test et la passation que vous souhaitez mettre à disposition des évalués. Une fois la passation sélectionnée, les informations qui y sont relatives apparaissent (l'intitulé, le type et la table de spécification).



## Paramétrage

Cette étape vous permet de paramétrer la passation de votre examen en sélectionnant notamment le nombre de formes parallèles de votre questionnaire ou le tarif de correction.

## Résumé

Le tableau présente l'ensemble des renseignements relatifs à la passation. Ce tableau se remplit au fur et à mesure de votre avancement dans les étapes de ce module.



## Information

Cette étape vous permet de lier une session à votre passation. Vous pouvez également y ajouter des remarques complémentaires et une description de la passation qui apparaîtront sur la première page de l'examen.



Pour confirmer l'encodage de vos informations, cliquez sur OK, sinon sur Annuler.

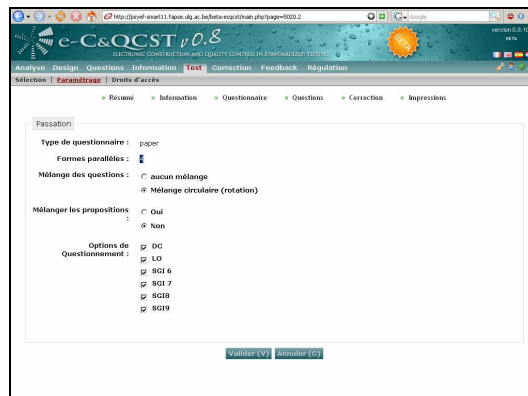
## Passation

A cette étape, vous choisissez le nombre des formes parallèles de questionnaires que vous souhaitez obtenir. Pour ce faire, il vous suffit d'encoder le chiffre correspondant dans la case prévue à cet effet.

Si vous souhaitez plusieurs formes parallèles, vous pouvez décider soit de mélanger les questions de manière aléatoire, ou de manière circulaire, soit de ne pas mélanger les questions.

Les propositions peuvent également être mélangées.

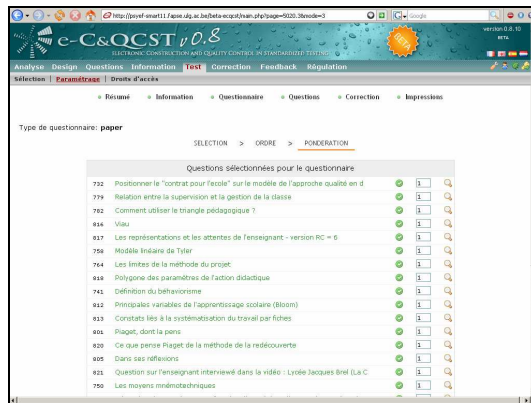
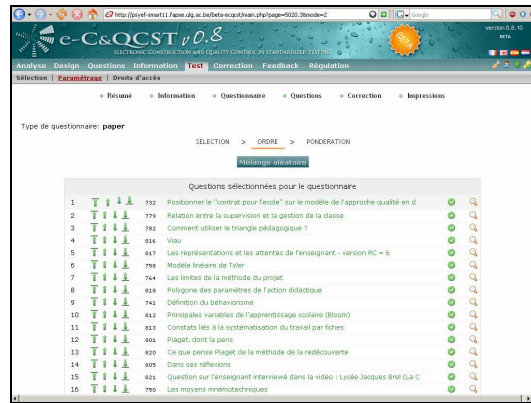
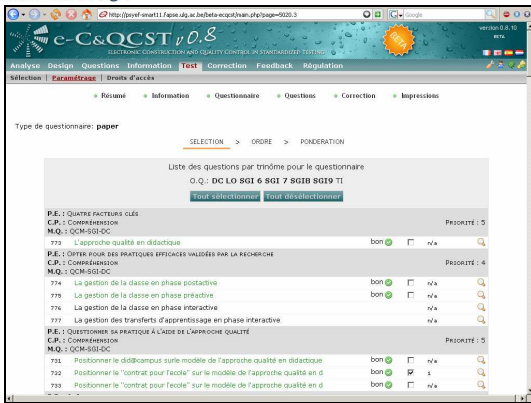
Vous pouvez également sélectionner des options de questionnement à apposer à votre passation.



Cliquer sur valider pour confirmer votre sélection, sinon sur annuler.

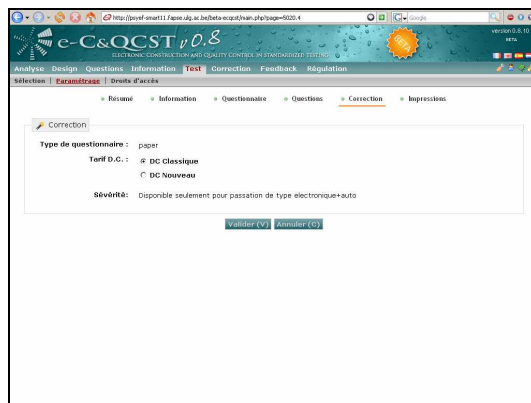
## Questions

A la sous-étape "sélection", la liste des questions encodées derrière les trinômes relatifs à votre passation apparaît. Vous pouvez sélectionner les questions que vous souhaitez intégrer dans votre examen. Vous passez ensuite directement à la seconde sous-étape intitulée "ordre". Vous pouvez y modifier l'ordre des questions présentées dans votre première forme de questionnaire. A la dernière sous-étape, vous pouvez associer une pondération aux questions qui feront partie de votre test.



## Correction

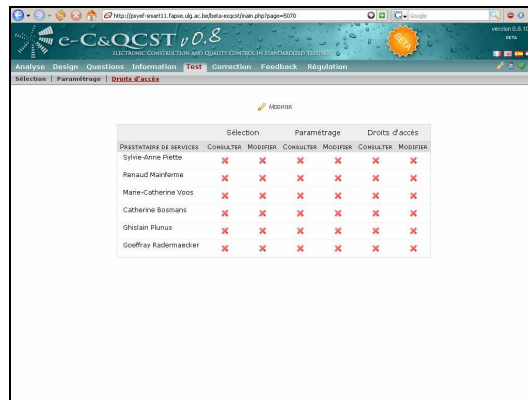
Cette étape vous permet de sélectionner le tarif de correction que vous souhaitez appliquer à votre passation. Des tarifs sont déjà proposés. Cependant, vous pouvez créer votre propre tarification. Pour ce faire, vous lui donnez un intitulé, déterminez les points accordés à une réponse correcte, les points déduits pour une réponse incorrecte ou pour une omission.



## Impression

Ce lien vous permet d'afficher des informations sur les questions du test - et notamment leur position dans l'évaluation.

Cette section permet d'accorder des droits (de lecture et/ou d'écriture) à des acteurs pour chacune des étapes présentes dans ce module.



## 2.6. Le module Correction

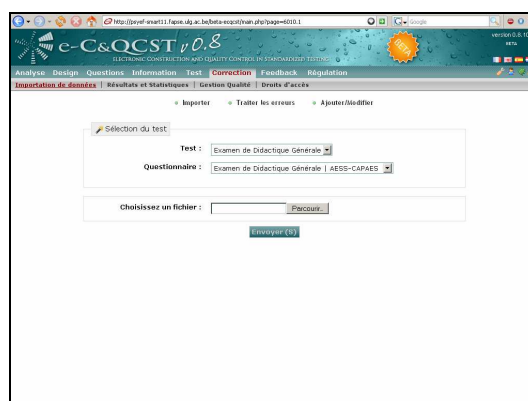
Ce module vous permet d'obtenir les résultats et des statistiques d'une passation que vous pourrez ensuite communiquer aux évalués.

### Importation des données

Au terme de cette étape, vous obtenez un tableau présentant le matricule des évalués qui ont réalisé la passation ainsi que leurs réponses.

### Importer

Cette étape vous permet d'importer les données relatives à une passation clôturée. Pour ce faire, sélectionnez le test et la passation concernés.



### Traiter les erreurs

A cette étape, vous pouvez traiter les erreurs éventuelles (dans le matricule ou les réponses) à partir du tableau présentant les matricules des évalués ainsi que leurs réponses.

## Ajouter/Modifier

Après avoir cliqué sur Ajouter/Modifier, vous pouvez ajouter une ligne relative à un évalué supplémentaire. Vous devez alors encoder son matricule, le numéro de son questionnaire ainsi que ses réponses.

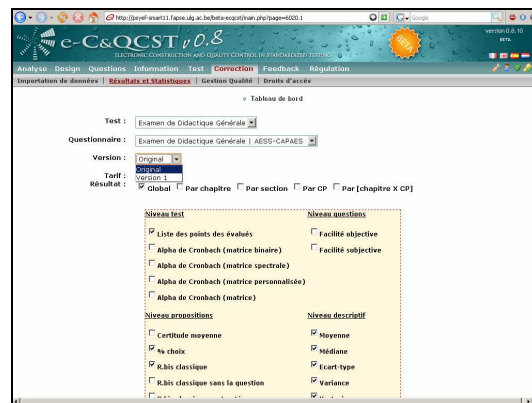
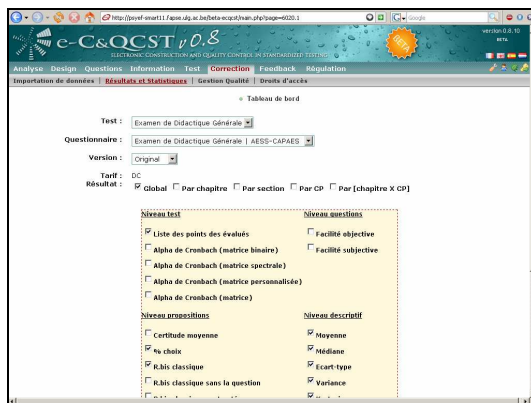
Cliquez ensuite sur Valider pour valider vos actions à cette sous-étape.

## Résultats et statistiques

A cette étape, vous obtenez les résultats relatifs à votre passation.

Certaines statistiques sont présélectionnées. Cependant, vous pouvez supprimer celles qui ne vous intéressent pas et en choisir d'autres.

Vous pouvez obtenir des statistiques au niveau de la globalité de l'examen, par chapitre, par section, par catégorie de performance, par binôme [catégorie de performance x CP].



Les statistiques requises sont alors présentées dans un document pdf ou html.

Pour visualiser les statistiques d'une nouvelle version (créée suite à des modifications réalisées à la sous-étape suivante "Gestion Qualité"), sélectionnez dans le menu déroulant le nom de la version concernée.

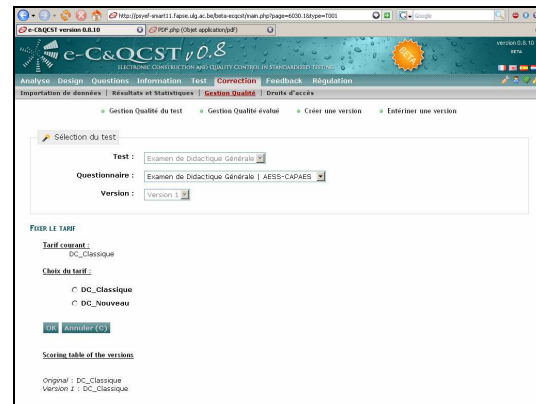
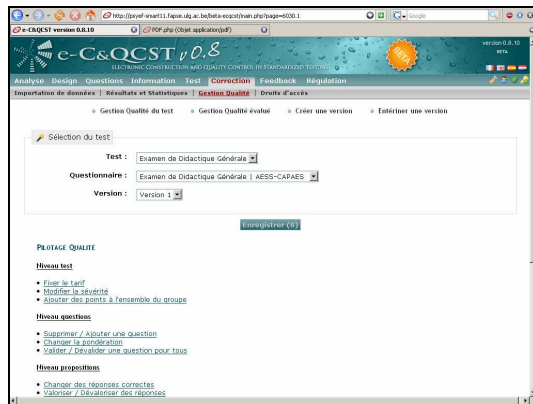
## Gestion Qualité

Cette étape vous permet de créer des versions supplémentaires liées à votre passation en améliorant la qualité au niveau du test ou au niveau de chacun des évalués. Elle permet également d'entériner une version afin ensuite de communiquer les résultats aux évalués.

### Gestion Qualité du test


A cette sous-étape, vous pouvez effectuer des modifications au niveau du test, des questions ainsi que des propositions. Le premier écran vous propose une série de paramètres liés à la correction du test. Sélectionnez une version créée. Les différents paramètres pouvant être modifiés deviennent alors sélectionnables. En cliquant sur le nom d'un paramètre, les changements possibles apparaissent. Il vous

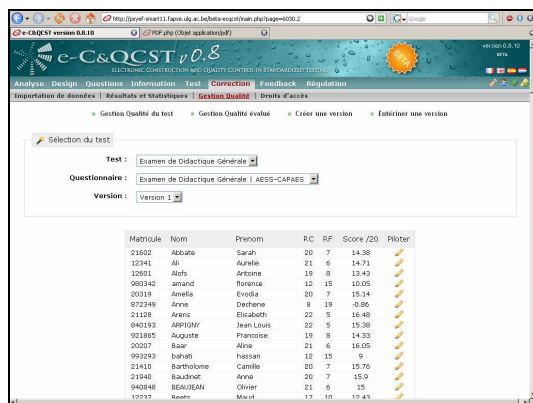
suffit ensuite de sélectionner l'élément souhaité et de confirmer votre choix en cliquant sur OK. Pour mémoire, les options choisies dans les versions précédentes vous sont rappelées.

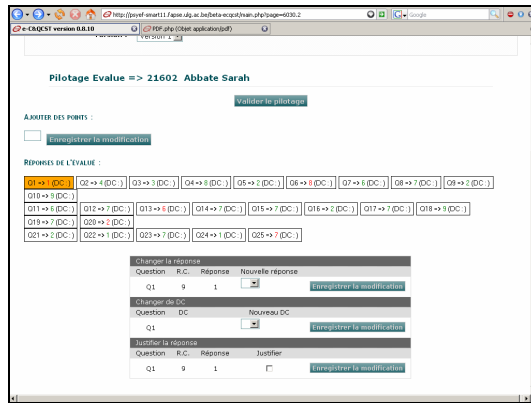


Vous pourrez obtenir les statistiques relatives à votre nouvelle version et ensuite les comparer avec celles de la version originale en sélectionnant votre version créée à l'étape "Résultats et Statistiques".

## Gestion Qualité évalué

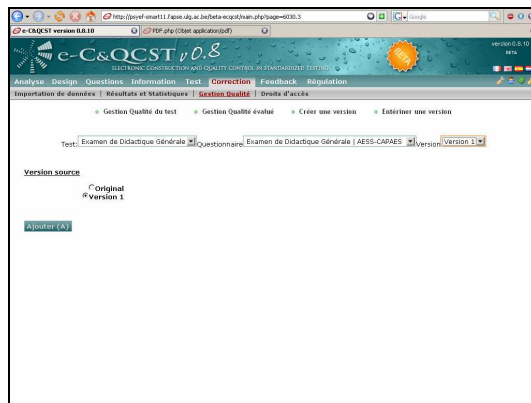
Cette étape vous permet de modifier les réponses encodées par les étudiants (par exemple, suite à la valorisation de certaines réponses correctement justifiées par les évalués). Les statistiques liées à chaque évalué sont fournies. En sélectionnant un évalué en cliquant sur l'icône , l'ensemble des réponses qu'il a données apparaît à l'écran. En cliquant sur une des questions, vous pouvez modifier la réponse encodée par l'évalué et/ou son degré de certitude.





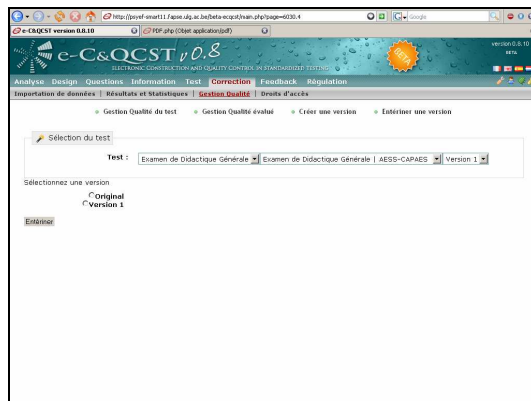
## Créer une version

En cliquant sur Ajouter, vous créez une nouvelle version de vos résultats et statistiques liés à une passation. Cette nouvelle version vous permettra d'effectuer des améliorations de la qualité au niveau du test et de l'évalué (cf. sous-étapes "Gestion Qualité du test" et "Gestion qualité évalué").

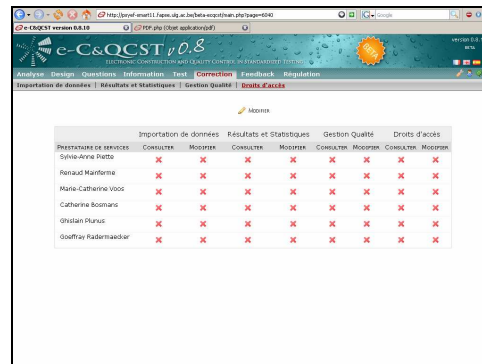


## Entériner une version

A cette étape, il vous suffit de sélectionner la version de vos résultats et statistiques que vous considérez comme la version finale. Les résultats de cette version seront communiqués aux évalués. Par la suite, vous ne pourrez plus modifier la version entérinée.



Cette section permet d'accorder des droits (de lecture et/ou d'écriture) à des acteurs choisis pour chaque étape présente dans le module.

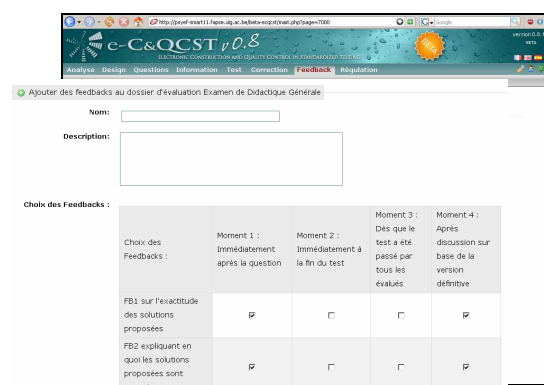


Personnes de services	Importation de données		Résultats et Statistiques		Gestion Qualité		Droits d'accès	
	Consulter	Modifier	Consulter	Modifier	Consulter	Modifier	Consulter	Modifier
Sylvie-Anne Piette	X	X	X	X	X	X	X	X
Renaud Marfème	X	X	X	X	X	X	X	X
Marie-Catherine Vooc	X	X	X	X	X	X	X	X
Catherine Boonans	X	X	X	X	X	X	X	X
Olivain Plunus	X	X	X	X	X	X	X	X
Geoffroy Rademaacker	X	X	X	X	X	X	X	X

## 2.7. Le module Feedbacks

Dans ce module, vous pourrez choisir ou modifier le type de feedbacks que vous souhaitez mettre à disposition des évalués. Vous pourrez leur communiquer divers rapports de feedbacks...

- Rapport de feedbacks sur l'exactitude des solutions proposées
- Rapport de feedbacks expliquant en quoi les solutions proposées sont correctes ou non
- Rapport de feedbacks partiel sur l'état des performances globales de l'évalué question après question
- Rapport de feedbacks sur les performances globales de l'évalué pour l'ensemble ou les différentes parties du test
- Rapport de feedbacks sur les performances du groupe pour l'ensemble ou les différentes parties du test
- Rapport de feedbacks sur la qualité des résultats aux propositions, questions et test



Ajouter des feedbacks au dossier d'évaluation Examen de Didactique Générale

Nom:

Description:

Choix des Feedbacks :

Choix des Feedbacks :	Moment 1 : Immédiatement après la question	Moment 2 : Immédiatement à la fin du test	Moment 3 : Dès que le test a été passé par tous les évalués	Moment 4 : Après discussion sur base de la version définitive
FB1 sur l'exactitude des solutions proposées	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
FB2 expliquant en quoi les solutions proposées sont correctes ou non	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>

Vous pourrez également choisir le moment de mise à disposition des feedbacks parmi les possibilités suivantes :

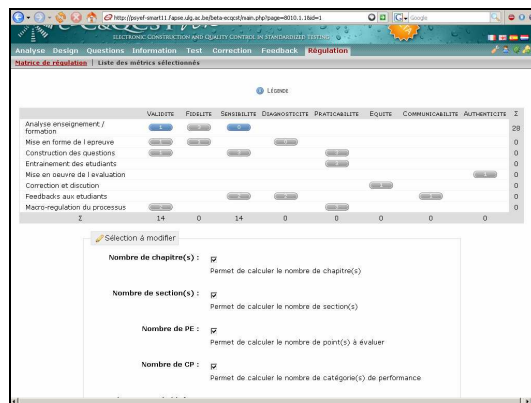
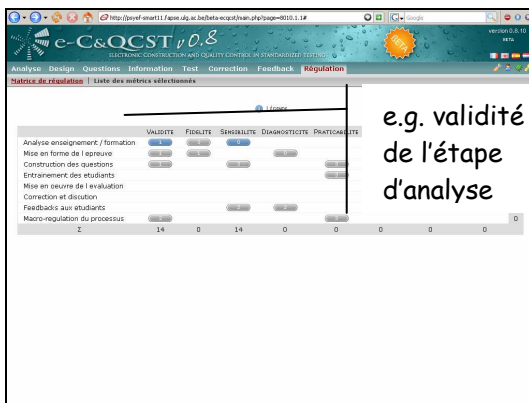
- Moment 1 : Immédiatement après la question
- Moment 2 : Immédiatement à la fin du test
- Moment 3 : Dès que le test a été passé par tous les évalués
- Moment 4 : Après discussion sur base de la version définitive

## 2.8. Le module de macrorégulation

Le module de macrorégulation vous permet d'analyser les performances de l'ensemble de votre test. Sur base de l'information ainsi fournie, vous pourrez en améliorer certains aspects pour une utilisation ultérieure.

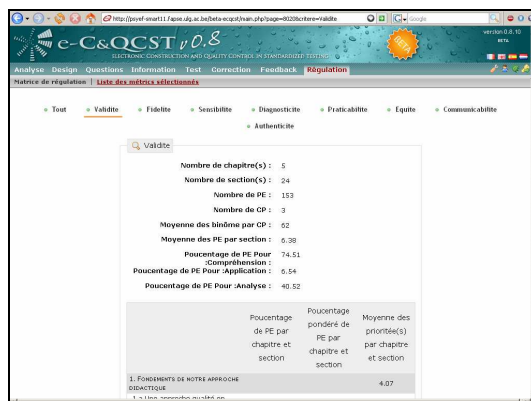
### Matrice de régulation

La première étape de la macrorégulation consiste à définir les thèmes que vous souhaitez approfondir lors de la macrorégulation. A cette fin, l'interface vous propose un tableau à double entrée croisant les différentes étapes du cycle d'une part et les critères de qualité attendus d'un test d'autre part. Sur chacun de ces croisements figure un indice (de 0 à 3) qui préfigure de l'importance de ce croisement pour la qualité du test. En cliquant sur l'un des croisements, l'interface propose une série d'indicateurs (métrics) en lien avec l'atteinte d'objectifs qualité liés à ce croisement. Vous pouvez alors sélectionner une série d'indicateurs.



### Liste des métriques sélectionnés

Lorsque des indicateurs ont été sélectionnés, leur liste est proposée afin d'être analysée. Par défaut, toutes les métriques sélectionnées seront affichées. Il est toutefois possible de les sélectionner en fonction des critères de qualité auxquelles elles se rapportent.



## 3. L'administration de la plateforme




Ce module "Administration", accessible via l'icône (toujours disponible en haut et à droite de l'interface) vous permet de créer vos épreuves, vos groupes d'évalués et vos taxonomies ainsi que de compléter les données relatives à votre profil.

### 3.1. Gestion des groupes d'évalués

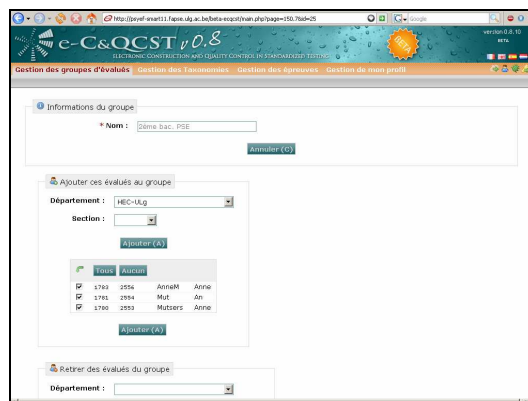
A cette étape, vous créez et gérez des groupes d'évalués.



En cliquant sur l'icône , vous ajoutez un groupe d'évalués auquel vous donnez un nom.

Ensuite vous pouvez gérer les groupes existants à partir des icônes suivantes :

 vous permet de visualiser les évalués appartenant à votre groupe ainsi que d'ajouter ou de supprimer un ou plusieurs évalués à ce groupe.



 vous permet de supprimer un groupe d'évalués.

 vous permet de modifier le nom d'un groupe d'évalués.

### 3.2. Gestion des taxonomies

Cette étape vous permet d'ajouter la ou les taxonomies que vous employez.

### 3.3. Gestion des épreuves

Au terme de cette étape, une ou plusieurs épreuves ont été créées et vous y avez associé un groupe d'évalués et des acteurs.



#### Afficher

Cette fonction vous permet d'afficher la liste des épreuves existantes.

#### Ajouter

Vous pouvez créer une nouvelle épreuve et y associer un groupe d'évalués.

#### Modifier

Cette fonction vous permet de modifier l'intitulé d'une épreuve existante ainsi que le groupe qui y est associé.

#### Supprimer

Vous pouvez supprimer une épreuve existante.

#### Gérer les évalués

Cette étape vous permet d'associer un groupe d'évalués à une épreuve.

#### Gérer les acteurs

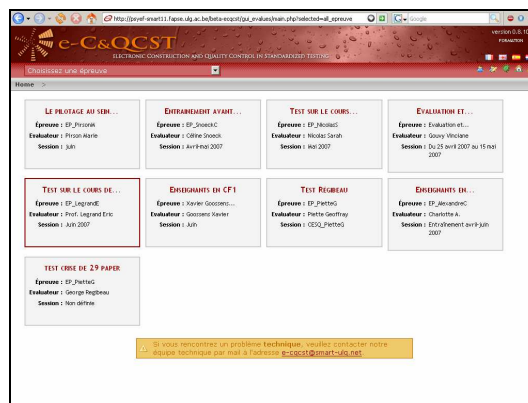
Vous pouvez associer des acteurs à votre épreuve afin de leur accorder des droits de lecture et/ou d'écriture. Ces acteurs apparaîtront à l'étape "gestion des acteurs" présente à la fin de chaque module.

### 3.4. Gestion de mon profil

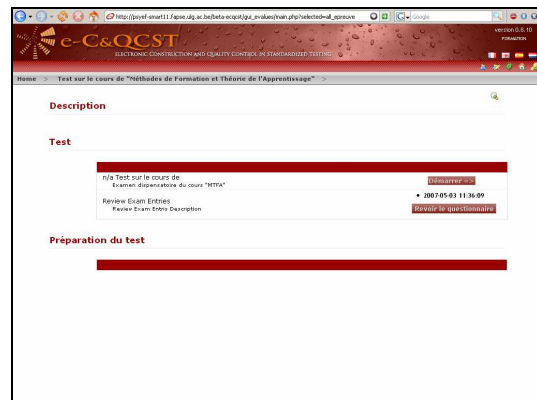
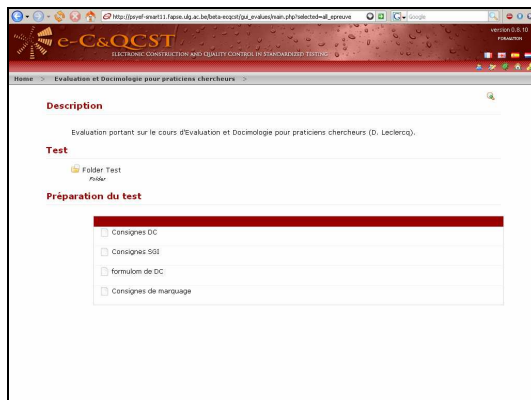
Cette étape vous permet d'associer, à votre nom d'utilisateur, des informations vous concernant.

### 4. Interface évalués

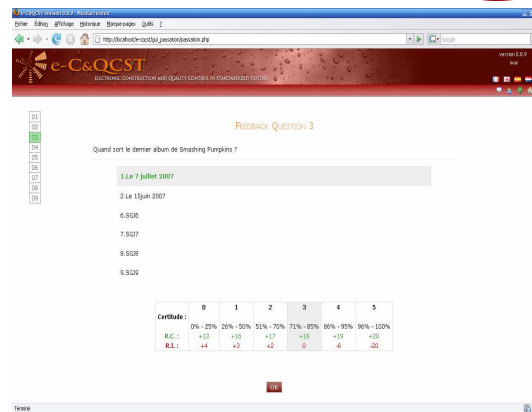
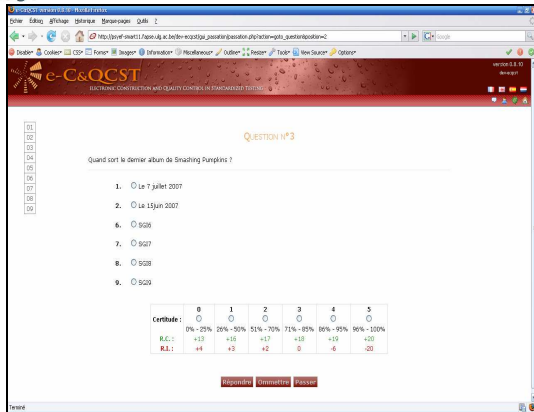
Après s'être enregistré à la plateforme, l'évalué a une vue d'ensemble des différents tests auxquels il est associé. Sur cette page d'accueil, il dispose d'informations globales relatives à chaque test (nom de l'épreuve, évaluateur,...). Par défaut, l'ensemble des tests auxquels il a accès sont affichés, mais il peuvent être affichés par épreuve suivant le menu « choisissez une épreuve »



En cliquant sur l'énoncé du test, l'évalué arrive sur une page dédiée au test concerné. A ce stade, trois cadres apparaissent. Le premier présente une description du test, le deuxième est relatif à la passation et le troisième expose les différents éléments liés à la préparation de cette passation.



Au niveau du cadre relatif à la passation, une fois la mise à disposition effective, vous avez accès au démarrage de la passation et aux documents qui y sont associés. Pour lancer la passation, l'évalué doit cliquer sur Démarrer la passation. L'évalué dispose alors d'une série d'informations concernant la passation, il peut lire les consignes, quitter cette page pour revenir à la page d'accueil ou démarrer réellement la passation. Lorsqu'il choisit cette option, la première question s'affiche



Lorsqu'il a répondu à la question un feedback lui est proposé si cette option a été choisie par le responsable d'évaluation.

## VII. CONCLUSIONS

### 1. Effets en retour du développement informatique sur nos conceptions docimologiques

Développer exAMS a eu un certain nombre d'effets en retour sur la modélisation théorique du Cycle de Construction et de Gestion Qualité des Tests Standardisés. Les éléments suivants expliquent les raisons de cet effet en retour.

#### a) Clarification et affinement des concepts liés au cycle CGQTS

Une première série d'effets en retour liés à la traduction informatique de nos conceptions docimologiques a été induit par l'impérieuse nécessité de formaliser dans le détail les concepts sous-jacents au cycle de Construction et de gestion qualité des tests standardisés (cycle CGQTS – voir supra) en vue de l'implémenter dans un outil informatisé. En effet, la pensée informatique demande un grand niveau de formalisation et un grand niveau de cohérence dans l'imbrication des différents concepts. Or, notre méthode, le cycle CGQTS, si elle s'appuyait sur de nombreux référents théoriques et avait elle-même fait l'objet de nombreuses publications, n'avait jamais été formalisée dans le niveau de détails requis pour son informatisation. Dès lors, un dialogue collaboratif soutenu avec les informaticiens a été nécessaire pour clarifier et préciser les concepts. La première étape a été de travailler à formaliser un langage commun : le langage scientifique a dû être compris et traité par l'équipe informatique et, inversement, les vocables et principes de programmation des informaticiens sont devenus familier aux chercheurs.

Dans ce contexte, la formalisation informatique du cycle CGQTS nous a amené à distinguer des concepts forts proches les uns des autres et qui avant ce travail collaboratif avec l'équipe informatique portaient le même nom. Par exemple, le concept de « test » que nous utilisions de manière globale et unifiée a dû être décomposé en « épreuve », « test », « dossier d'évaluation » et « passation ».

Outre cette clarification conceptuelle, la programmation du cycle CGQTS a soulevé des questions épineuses du point de vue scientifique. Par exemple, en travaillant sur le module de création des questions, nous avons conçu la plate-forme exAMS de manière à ce qu'une question soit toujours attachée à un et un seul trinôme (PE x CP x MQ – voir supra), ce qui est vrai dans la plupart des cas de figure que nous sommes amenés à traiter classiquement dans nos tests. Or, il existe des situations docimologiques où des questions peuvent couvrir différents points à évaluer. Quand la table de spécification et les questions étaient rédigées sur support papier, une note en bas de page suffisait pour expliquer cette transversalité. Via exAMS, la formalisation du processus est telle qu'une simple annotation n'est plus possible au regard de l'impact sur les procédures informatisées, ce sont de nouvelles formalisations et de nouvelles procédures qui doivent alors être trouvées.

En conclusion de ce premier effet en retour, si nos informaticiens ont souvent dû trouver des solutions techniques à certains problèmes scientifiques, les chercheurs de notre équipe ont aussi été amenés à trouver des solutions scientifiques à des problèmes conceptuels mis en lumière par la formalisation informatique.

**b) La nécessaire formalisation et gestion informatisée des écarts entre tâches docimologiques prescrites et tâches docimologiques effectuées**

Un deuxième effet en retour du développement informatique sur nos conceptions docimologiques a été provoqué par la nécessité de préciser le paramétrage de nos procédures. Avant de disposer de la plate-forme exAMS, une série d'actions faites par l'opérateur humain qui mettait en œuvre la méthodologie de la construction et gestion qualité des tests standardisés étaient implicites et/ou globalement définies. Ainsi, cela laissait au docimologue l'opportunité d'adapter les tâches prescrites dans la méthodologie en tâches effectives qui répondent au mieux au contexte et aux objectifs poursuivis.

Cette distinction entre tâche prescrite et tâche effective est intéressante. En ergonomie, on distingue la tâche prescrite, celle qui est donnée par les prescripteurs de la tâche effective, de celle que le sujet effectue réellement lorsqu'il est dans une situation précise de travail.

Le processus de transposition d'une tâche donnée en tâche effective, chez un évaluateur est le suivant. Tout d'abord, ce dernier récupère, de son prescripteur (le modèle du cycle CGQTS) un certain nombre de consignes, de prescriptions qui lui permettent de se représenter un premier état de sa tâche, nommée tâche prescrite. Cette tâche prescrite va être ensuite redéfinie, en fonction d'une situation précise, puis, une nouvelle fois transformée, ici dans l'activité de création d'un test, en prenant en compte les caractéristiques les plus immédiates de la situation : c'est la tâche effective, celle qui est véritablement mise en œuvre dans l'action. *A posteriori*, l'agent met à jour une nouvelle représentation de la tâche, celle de la tâche réalisée.

Notons que chacune des étapes précédentes donne lieu à une transformation, faible ou plus importante, de la tâche. L'importance de ces transformations dépend d'un grand nombre de facteurs : la précision des prescriptions, l'expertise de l'agent, les contraintes de la situation, etc. Il convient de souligner que ces écarts ne sont pas nécessairement des erreurs ou des dysfonctionnements de la part de l'évaluateur : d'une part, la tâche docimologique prescrite contient toujours des implicites que l'évaluateur doit régler, d'autre part, ces écarts peuvent être dus à une nécessaire adaptation de l'évaluateur aux contraintes du contexte de l'évaluation.

Ainsi, avant exAMS, quand la tâche était gérée par un humain, les aspects implicites et les aspects de contextualisation étaient du ressort de l'opérateur humain. A partir du moment où ces tâches sont automatisées ou semi automatisées à travers la plate-forme exAMS, l'adaptation au contexte a dû être pensée et définie à l'intérieur même de nos procédures. Chaque cas de figure a donc dû être compris, mesuré, pondéré et des décisions associant certains aspects scientifiques ont dû être prises pour intégrer l'ensemble des possibles au niveau de notre plate-forme.

**c) De nouvelles possibilités qui stimulent la pensée créative du chercheur en docimologie**

Un troisième retour du développement sur nos conceptions en évaluation a trait aux perspectives futures en termes de recherches docimologiques qui sont potentiellement offertes par l'usage de l'outil exAMS. Ainsi, l'ensemble des données stockées dans la plate-forme permettra de répondre assez aisément à des questions de recherche qui, sans cela, aurait nécessité la mise en œuvre de méthodologies complexes. A partir des données stockées dans la plate-forme nous pourrions étudier les conditions qui augmentent l'efficacité docimologique des épreuves. Par

exemple il sera beaucoup plus aisé d'étudier l'impact des entraînements sur l'évolution des performances des évalués lors des tests certificatifs. Les résultats de ce type de recherche devraient nous permettre de préciser, voire de recommander, des paramétrages d'entraînement en fonction des contextes des évaluations. Bref, de nouvelles questions de recherche se profilent déjà en lien avec l'usage d'exAMS et l'analyse des données docimologiques que cette plate-forme va permettre d'enregistrer.

## 2. Perspectives futures

### a) Adapter la plate-forme exAMS à l'évaluation et à la gestion des compétences

Les objectifs de ce projet ont ainsi été atteints : nous avons mis au point un « *examination Assessment Management System (exAMS)* » permettant de concevoir, construire, mettre en œuvre, traiter et corriger des évaluations standardisées pour le monde de l'enseignement et de la formation. Les procédures qui sont mises en œuvre respectent la méthodologie du cycle CGQTS et permettent d'atteindre de hauts niveaux de qualité selon les critères définis dans notre modèle docimologique.

Parallèlement, depuis plusieurs années le SMART, et plus récemment sa spin-off Assess Group S.A., proposent leur approche qualité en évaluation dans le cadre d'activités de consultance. Dans ce contexte, de grandes firmes telles que FORTIS BANK, AXA et de grandes administrations comme l'Institut de Formation des Administrations fédérales (IFA) et le FOREM font appel déjà à nos services et utilisent déjà, pour certains, notre plate-forme « exAMS ». Nos contacts avec ces entreprises au travers des prestations de services fournis par notre spin-off Assess Group S.A., nos interactions avec les partenaires de ce projet e-C&QCST, ainsi que notre expertise scientifique, nous amènent à la conclusion qu'il existe un large marché pour un outil qui permettrait d'évaluer de façon automatisée ou semi automatisée des performances complexes liées à des ensembles intégrés et fonctionnels de savoirs, savoir-faire et savoir être en jeu dans des catégories de situations où les personnes doivent faire la preuve de leurs compétences.

Par exemple, dans le cadre de nos prestations docimologiques au sein de ces grandes entreprises et institutions, nous avons constaté l'attrait du marché pour de nouvelles ressources logicielles qui permettraient d'évaluer, de manière authentique, automatisée et objective des compétences de haut niveau.

Le monde de l'éducation n'échappe d'ailleurs pas à l'approche par compétences. Certaines réformes de grande ampleur, ont été mises en œuvre dans ce sens dans plusieurs systèmes scolaires. Par exemple, au Canada la réforme des programmes du Ministère de l'Éducation du Québec, initiée dans les années nonante, propose de prendre en compte les compétences nouvelles qu'exigent les nouveaux systèmes de travail et les nouvelles situations d'emploi. En Belgique, le parlement de la Communauté française a voté en 1999 le décret fixant les « socles de compétences ».

En corollaire, un renouveau scientifique en évaluation émerge depuis les années nonante sous l'impulsion des approches par compétences qui remettent à l'avant plan la nécessité d'une évaluation instrumentée des performances complexes. Ce nouveau courant de recherches docimologiques ouvre de nouvelles perspectives très prometteuses en matière d'évaluation tout en pointant d'énormes problèmes en matière de praticabilité ainsi qu'au niveau de la maîtrise de phénomènes qui entachent la fidélité de telles évaluations : concordance intra-juge et inter-juges,

effets d'interaction, ... (e.g. Shavelson & al., 1993; Suh, 2001; Berry & Mielke, 2002; Gilles, 2002; Scallon, 2004).

On le voit, les enjeux sont énormes pour la *R&D* en termes scientifiques, économiques et sociaux. Du point de vue des ressources humaines ou de l'employabilité, établir des bilans de compétences fiables est un enjeu majeur non seulement pour le développement économique des entreprises mais aussi pour la lutte contre le chômage et les inégalités sociales.

Notre approche scientifique et notre connaissance du marché nous permettent d'affirmer que le SMART dispose du *know how* en ingénierie docimologique lui permettant de relever les défis d'une recherche ayant pour but d'instrumenter l'évaluation des compétences à l'aide de méthodes et de procédures de qualité.

Dès lors, nous envisageons d'enrichir la plate-forme exAMS par de nouvelles fonctionnalités d'évaluation des performances complexes dans le contexte des évaluations de compétences de niveau élevé, et ce, dans des situations plus authentiques, plus proches de la vie courante. Cet enrichissement nécessitera de concevoir de nouvelles procédures docimologiques pour le modèle en cycle de construction et de gestion qualité utilisé par la plate-forme exAMS et d'y greffer informatiquement ces nouvelles fonctionnalités.

Nous pensons que ces nouvelles fonctionnalités ajoutées à notre plate-forme apporteront une réelle plus value au paysage de la gestion de compétences.

En effet, augmentant notre plate-forme exAMS par de nouvelles fonctionnalités capables d'instrumenter un éventail plus large de techniques docimologiques, incluant des outils de mesure traditionnels, par exemple les questions à réponses ouvertes longues, et de nouvelles pratiques évaluatives issues du renouveau en évaluation, par exemple les portfolios de compétences, (Howe & Ménard, 1993; Biggs, 1995; Ryan & Myasaka, 1995; Scallon, 1996; Legendre, 2001; Scallon, 2004) et en puisant ces techniques dans les fondements de la méthodologie d'observation, de jugement de performances ou d'évaluation de projets (Forcier, 1991; Linn & Grondlund, 2000; Popham, 2000; Nitko, 1996), nous augmenterons considérablement la puissance de notre plate-forme exAMS.

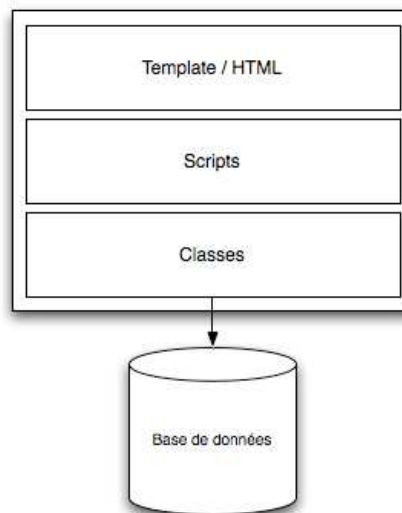
Les développements futurs passeront donc par des *R&D* visant la création de procédures nouvelles à greffer sur exAMS et qui devront permettre à chaque évalué :

- (1) de construire des réponses élaborées dans le cadre de situations de performances complexes;
- (2) d'être confronté à des situations reflétant le réel et qui ont du sens pour lui;
- (3) d'interagir avec le système d'évaluation, notamment d'aller vers l'information en vue d'offrir des solutions originales;
- (4) d'être jugé sur sa performance selon différents aspects;
- (5) de bénéficier des apports d'une autoévaluation et d'une co-évaluation de ses performances;
- (6) d'obtenir des garanties en ce qui concerne la validité et la fidélité des jugements portés sur ses performances;
- (7) d'être évalué non seulement sur le produit de sa performance, mais aussi sur le processus qu'il a mis en œuvre;
- (8) d'éventuellement être jugé dans le cadre de travaux effectués en groupe;
- (9) d'obtenir des feedbacks qui lui permettront d'élever son niveau de compétence.

### b) Vers une architecture permettant encore plus d'interopérabilité ainsi qu'un travail collaboratif avec une communauté de développeurs

Dès le début, le projet e-C&QCST a été orienté dans une logique dite « *open source* », ceci afin de permettre à un large public de l'utiliser mais aussi de s'approprier l'application de manière à pouvoir l'adapter aux besoins spécifiques des utilisateurs. Dans cette optique, le choix d'une plateforme de développement ouverte et gratuite s'est imposé comme une évidence.

Nous avons donc mis en place un projet en utilisant les méthodes classiques appliquées aux développements basés sur les technologies apache/PHP/MySQL.



Après plusieurs mois de développement, le code de l'application devenant de plus en plus complexe, une série de problèmes sont apparus qui ont remis en cause notre conception initiale du développement collaboratif :

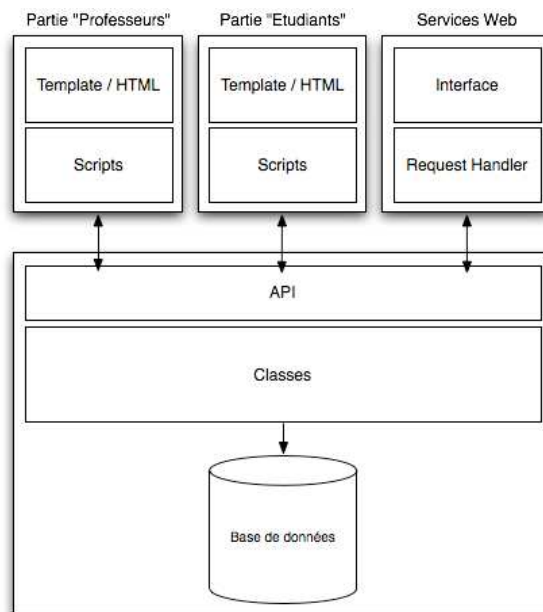
1. L'ensemble du code représente un nombre important de lignes et il ne serait par conséquent pas simple pour un néophyte de s'y plonger ;
2. Certaines méthodes ou procédures implémentent des comportements qui ne devraient pas d'un point de vue documentaire être modifiés sans une réflexion solide et une analyse précise de l'impact de ces changements sur la méthodologie proposée ;
3. La compatibilité des développements de l'équipe du SMART avec ceux des développeurs tiers pourrait poser problème, notamment lors de la mise à disposition de nouvelles versions de notre application.

Dès lors, nous avons modifié notre manière d'aborder la mise à disposition de notre application pour faciliter le processus d'appropriation par des entités externes et assurer une pérennité des développements basés sur notre plateforme exAMS.

Ce changement d'approche a modifié en profondeur la structure de nos développements et nous a conduit à réorganiser les différentes couches logicielles d'exAMS en vue d'offrir une plate-

forme sur laquelle un développement tiers pourra venir se greffer, tirer parti des services rendus par nos outils et assurer une compatibilité avec les futurs développements.

Cette migration menée au travers d'un processus de mise en place d'une API (Application Programmer Interface) permet d'assurer une indépendance entre une application spécifique et notre modèle de données, elle est le cœur de notre plate-forme exAMS. L'API améliore la gestion des différents éléments de notre application et autorise une meilleure gestion de leurs spécificités. Par exemple, nous avons scindé notre application en deux parties distinctes : une première dédiée à la création de tests standardisés, leur correction, etc. et la seconde dédiée aux évalués et à la passation de tests en ligne. Nous avons pu ainsi fournir des interfaces plus claires et plus facilement exploitables.



Nous voulons également offrir une plus grande interopérabilité via une interface dédiée aux WebServices. Ainsi, une application tierce existante peut assez simplement interroger notre plate-forme, en recevoir certaines informations et même lui fournir une collection de données qui seront intégrées et traitées par exAMS.

Par exemple, nous avons réalisé un prototype qui nous permet de récupérer l'ensemble des données relatives à un test : liste des étudiants, des questions, des réponses correctes, des réponses fournies par les étudiants, certains paramètres de correction, etc. Ces données sont disponibles dans une application tierce (TVI - voir supra) via une requête sur nos services Web et nous sommes en mesure de valider l'ensemble de ces données, d'insérer dans nos bases de données toutes les informations transmises, d'effectuer une correction du test et de fournir un résultat complet au format PDF.

Dès lors, des interactions sont désormais envisageables avec d'autres applications. Par exemple, il est possible :

- d'intégrer une table de spécification (TDS) exAMS dans un Learning Management System (WebCT, Dokeos, Moodle, etc.) et d'associer aux questions liées à cette TDS des informations,

- notamment de remédiation (ce qui permet d'enrichir considérablement les fonctionnalités de notre plate-forme exAMS, mais aussi des classiques LMS) ;
- l'insertion automatique des résultats obtenus après correction avec la plate-forme exAMS dans un système de gestion de fiches d'étudiants comme par exemple le système Penelope utilisé à l'ULg ;
  - l'intégration des informations d'exAMS avec un outil GRH pour concevoir des parcours d'apprentissage/évaluation spécifiques à une personne, un groupe ;
  - ...

Dans le futur, l'équipe informatique continuera à travailler sur l'API afin de la rendre encore plus complète et d'en simplifier encore plus l'usage. Ainsi, il sera possible de concevoir des applications dédiées interopérables avec notre plate-forme exAMS. Par exemple, un module d'encodage de questions qui répondraient à des besoins très précis exprimés par un centre de formation ou un module de gestion des évalués qui soit directement relié à la base de données du personnel d'une entreprise ; etc.

### 3. Conclusion générale

On le voit, que ce soit du point de vue de la recherche docimologique ou du point de vue du développement informatique, le travail qui a été réalisé dans le cadre de cette R&D « *electronic Construction and Quality Control in Standardized Testing (e-C&QCST)* » est impressionnant. Un aspect remarquable, qu'il convient de souligner encore une fois, réside dans le fait que la plate-forme exAMS, est actuellement déjà utilisée par des acteurs économiques, dont l'important organisme bancaire AXA. Mais le plus remarquable selon nous c'est que nous venons de créer un des premiers AMS, et à notre connaissance le tout premier qui soit « *open source* ». Avec la plate-forme exAMS, nous sommes aujourd'hui en pointe dans le domaine de l'ingénierie docimologique et parmi les pionniers des développeurs d'une nouvelle lignée d'applications informatiques à ranger sous le vocable « *Assessment Management System (AMS)* ».

Enfin, rappelons qu'une spin off intitulée Assess Group S.A. a récemment été créé (en octobre 2006) par le SMART avec l'aide du programme First spin off de la DGTRE. Cette spin off qui emploie déjà 6 personnes dispose maintenant de la plate-forme qui va lui permettre de déployer de nouvelles offres de services dans le domaine de l'ingénierie docimologique et de s'ouvrir à de nouveaux marchés. La valorisation des résultats de cette R&D participe ainsi à l'émergence d'un nouveau pôle d'excellence docimologique en Région wallonne.

Aujourd'hui, cette excellence en ingénierie docimologique de nos équipes est déjà reconnue à l'international. Par exemple, Assess Group S.A. a remporté un marché public européen en vue de l'organisation et de la mise en œuvre d'un concours de recrutement pour un ministère de la République française. Dans ce contexte précis, la plate-forme exAMS est également utilisée par Assess Group S.A.

Grâce à la R&D « *electronic Construction and Quality Control in Standardized Testing (e-C&QCST)* » et à la plate-forme exAMS qui en est issue, les membres de notre équipe de recherche et de notre spin off ont acquis aujourd'hui une grosse avance dans la maîtrise des processus de

construction et de gestion instrumentée de la qualité des tests standardisés. Assess Group S.A. dispose maintenant d'arguments de poids qui la distingue de ses concurrents et l'équipe du SMART a consolidé son « *know how* » en ingénierie docimologique. Et ce, dans un contexte où l'avenir est très ouvert, tant du point de vue de la recherche que de la valorisation économique car les avancées scientifiques issues de cette R&D répondent à de réels besoins sociétaux exprimés par des acteurs socio-économiques importants.

Nous tenons à terminer ce rapport en remerciant la DGTRE et la Région wallonne pour le soutien qu'elle nous a offert tout au long de ce projet de recherche du programme WIST. Nous remercions également la DGTRE et la Région wallonne d'avoir permis en marge de cette R&D, via le programme First spin off, la constitution d'une entreprise de valorisation de services en ingénierie docimologique qui participe au redéploiement économique de notre région.

## VIII. BIBLIOGRAPHIE

- Adams, J.-K. & Adams, P.-A., (1961). Realism of confidence judgments, *Psychological Review*, 68, 33-45.
- Agazzi, A. (1967). *Les aspects pédagogiques des examens*, Strasbourg : Conseil de l'Europe, C.C.C.
- Ahlgren, A. (1967). *Confidence on Achievement Tests and the Prediction of Retention*. Harvard University : Thèse de doctorat.
- Albert, M. & Raiffa, H. (1982). A progress report on the training of probability assessors. In Kahneman, D., Slovic, P. & A. Tversky A. (Eds), *Judgment under uncertainty : Heuristics and biases* (294-305), Cambridge : University Press, 1982.
- Baker J.-D. (1969). The uncertain student and the understanding computer, *La Recherche en enseignement programmé*, Tendances actuelles, Paris : Dunod, 303-319.
- Barnabé, C. (1995). *Introduction à la qualité totale en éducation*. Québec : Presses Inter Universitaires.
- Barrows, H.-S. & Tamblyn, R.-M. (1977). The portable patient problem pack (P4), a problem based learning unit, *Journal of Medical Education*, 52, 1002-1004.
- Barrows, H.-S. & Tamblyn, R.-M. (1980). *Problem based learning, an approach to medical education*. New York : Springer Publishing.
- Blais, J.-G., Laurier, M., Van der Maren, J.-M., Gervais, C., Lévesque & M., Pelletier, G. (1997). *L'évaluation des apprentissages à l'Université de Montréal et dans ses écoles affiliées*. Montréal : Université de Montréal, Faculté des sciences de l'éducation, Groupe de Recherche Interdisciplinaire en Pédagogie Universitaire (GRIPU).
- Bloom, B. (1969). *Taxonomie des objectifs pédagogiques - I. Domaine cognitif*, traduit par M. Lavallée. Montréal : Education Nouvelle.
- Bloom, B., Engelhart, M.-D., Forst, E.-J., Hill, W.-H. & Krathwohl, D.-R. (1956). *Taxonomy of educational objectives : handbook I, cognitive domaine*. New York : D. Mac Kay.
- Bonniol, J.-J. (1972). *Les comportements d'estimation dans une tâche d'évaluation d'épreuves scolaires - Etude de quelques-uns de leurs déterminants*. Aix-En-Provence : Université de Provence.
- Boucher, S. et al. (1997). *L'évaluation de la qualité de l'enseignement dans les institutions universitaires de la Communauté française de Belgique*, groupe de travail du CreF (Conseil des recteurs Francophones).
- Boxus, E. (1981). Une tentative d'insertion de l'évaluation formative dans la pratique quotidienne : l'opération PREDIC, in *Revue de l'Organisation des Etudes*, Bruxelles, 9, 9-39.
- Boxus, E. & al. (1991). *Principes communs pour évaluer les résultats cognitifs de la formation*. Bruxelles : Commissions des Communautés européennes, programme Eurotecnet.
- Brennan, R.-L. (1972). A generalized upper-lower item discrimination index. *Educational and psychological measurement*, 32, 289-303.
- Brier, G.W. (1950). Verification of forecasts expressed in terms of probability, *Monthly Weather Review*, 75, 1-3.
- Brisbois R. (1967). *Les corrélations en pédagogie et en psychologie*. Montréal, Canada : Les éditions Lidec.
- Bruno J. (1990). Confidence contour tests item analysis with information referenced testing. *Proceedings of the Seventh International Conference on Technology and Education*, Brussels.
- Bruno, J. & Baxter, J. (1989). An application of information reference testing, *Proceedings of the Sixth International Conference on Technology and Education*, Orlando, vol. 2, 191-192.

- Bruno, J. (1986). Assessing the Knowledge base of students: An information theoretic approach to testing. *Measurement and evaluation in counseling and development*, 19(3), 116-130.
- Bruno, J. (1987). Admissible probability measurement in instructional management. *Journal of computer based instruction*, 14(1), 23-30.
- Cardinet, J., Tourneur, Y. et Allal, L. (1975). Extension of generalisability theory and its applications in educational measurement, *Journal of Educational Measurement*, 18, 183-204.
- Carrall, J. M. (1995). *Scenario-Based Design*. New York : Willey.
- Carver, R.-P. (1974). Two dimensions of tests : Psychometric and Edumetric, *American Psychologist*, 29, 512-518.
- Castaigne, J.-L. & Gilles, J.-L. (2000). Application du cycle gestion qualité SMART des tests pédagogiques dans le cours d'Obstétrique et de Pathologie de la Reproduction des ruminants, équidés et porcins, Communication présentée lors des journées Form@sup, Liège : Université de Liège, Service de Technologie de l'Education.
- Castaigne, J.-L., Gilles, J.-L. & 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.
- Chicago Board of Education (1999). *Standardized Test Preparation – Preparing Your High School Students to take Standardized Tests*, Chicago : Chicago Public Schools.
- Constantine, L. L. (1996a). « Re : Architecture. » *Software Development* 4 (1).
- Constantine, L. L., Lockwood L. A. (1999) "Software for use", ACM Press, New York.
- Cox, R.C. & Vargas, J.S. (1966). *A comparison of item selection techniques for norm-referenced and criterion-referenced tests*. Paper presented at the annual meeting of the National Council on Measurement in Education.
- Cronbach, L. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16.
- Cronbach, L., Gleser, G.C. & Rajaratnam, N. (1963). Theory of generalizability. A liberalization of reliability theory. *British journal of mathematical and statistical psychology*, 16, 137-173.
- Cross, L. & Frary (1977). An empirical test of Lord's theoretical results regarding formula scoring of multiple choice tests, *Journal of Educational Measurement*, vol 14, 313-321.
- Dagnélie, P. (1992). *Statistique Théorique et Appliquée – Tome I*. Gembloux : Les presses agronomiques de Gembloux.
- D'Hainaut, L. (1973). Etude d'une nouvelle variable pour l'analyse statistique des expériences pédagogiques, *Bulletin de psychologie*, 305, vol. 26, 622-630.
- D'Hainaut, L. (1975). *Arrondir et estimer, cours programmé*. Paris : Hachette.
- De Bal, R., De Landsheere, G., & Beckers, J. (1977). *Construire des échelles d'évaluation descriptives*. Bruxelles : Ministère de l'Education, Organisation des Etudes.
- De Finetti, B. (1965). Methods for discriminating levels of partial knowledge concerning a test item, *British Journal of Mathematical and Statistical Psychology*, 18, 87-123.
- De Landsheere (1980). *Evaluation continue et examens - Précis de docimologie*. Bruxelles : Ed. Labor, Education 2000.
- De Landsheere, G. (1976). *La formation des enseignants demain*, Collection E3, Orientation, Ed. Casterman.
- De Landsheere, G. (1979). *Dictionnaire de l'évaluation et de la recherche en éducation*. Paris : Presses Universitaires de France.

- De Landsheere, V. (1986). *Faire réussir, faire échouer. La compétence minimale et son évaluation*. Paris : PUF.
- Debry, M., Deltour, J.-J., Demeuse, M., Tanguy, E., Gilles, J.-L., Leclercq, D., Malherbe, E., Perée, F., & Poncin, P. (1999). *Observations sur les Etudes Universitaires à la Faculté (CEUF) – Rapport de synthèse de la Commission Premier Cycle*. Liège : Université de Liège, Faculté de Psychologie et des Sciences de l'Éducation.
- Debry, M., Leclercq, D. & Boxus, E. (1998). De nouveaux défis pour la pédagogie universitaire. In D. Leclercq (Ed), *Pour une pédagogie universitaire de qualité*. Liège Sprimont : Mardaga.
- Deming, W.E. (1992). *The New Economics for Industry, Government, Education*. Cambridge MA : MIT Center for Advanced Engineering Study.
- Descartes, R. (1628). *Règles pour la conduite de l'esprit*.
- Descartes, R. (1636). *Discours de la méthode pour bien conduire sa raison et chercher la vérité dans les sciences*.
- Didier, P., Faivre, M., Fauroux, R., Grandpierre, A., Millord, D. & Rousselet, M. (1966). *Le bouton du mandarin. L'école face à notre avenir*. Casterman, Centre d'Études Pédagogiques.
- Dirkzwager, A. (1993). computer environment to develop valid and realistic predictions and self-assessment of knowledge with personal probabilities, in D. Leclercq & J. Bruno (Eds), NATO ASI Series, *Item Banking: Interactive Testing and Self Assessment*. Berlin: Springer Verlag, Vol. 112, pp. 146-166.
- D'Ivernois J.-F., Gagnayre R. (1995). *Apprendre à éduquer le patient*. Paris : Vigot.
- Droesbeke, J.-J., (2001). *Éléments de statistique*. Bruxelles : Editions de l'Université de Bruxelles.
- Dupont, P. & Ossandon, M. (1994). *La pédagogie universitaire*. Paris : Presses Universitaires de France.
- Ebel, R.L. (1965). Confidence–Weighting and Test Reliability, *Journal of Educational Measurement*, 2, 49-57 B.
- Ebel, R.L. (1968). Review of Valid confidence testing demonstration kit, *Journal of Educational Measurement*, 5, 353-354.
- Ebel, R.L. (1979). Using tests to improve learning, *Arithmetic-Teacher*, v27, n3, p10-12.
- Edwards, W. (1967). Probabilistic information processing by men and man-machine systems, in *La simulation du comportement humain*. Paris : Dunod, p. 187.
- Findley, W.G. (1956). A rationale for evaluation of item discrimination statistics. *Educational and Psychological Measurement*, 16, 175-180.
- Georges, F., Gilles, J.-L., Pirson, M., Simon, F. & Leclercq, D. (2001). Les feedbacks aux étudiants et aux enseignants du projet MOHICAN, in Leclercq, D. (Ed.), *Le premier des MOHICANs – Une recherche-action de Monitoring Historique des CANDidatures*, Rapport relatif au contrat 798363 entre le Conseil Inter Universitaire Francophone (CIUF) et l'Université de Liège. Bruxelles : CIUF (à paraître).
- Gibbs, G. Jenkins, A. & al. (1992). *Teaching large classes in higher education - How to maintain quality with reduced resources*. London : Kogan Page.
- Gilles, J.-L. (1995). Entraînement à l'autoévaluation : une comparaison filles/garçons à l'université, *Actes du Colloque de l'AIPU « Enseignement supérieur : stratégies d'enseignement appropriées »*, Hull : Université du Québec à Hull, pp. 159-166.
- Gilles, J.-L. (1996a). Utilisation des degrés de certitude et normes de réalisme en situation d'examen et d'auto-estimation à FA.P.S.E. - ULG, Communication présentée au *Colloque de l'Association pour le Développement des Méthodologies d'Évaluation en Éducation (ADMEE-EUROPE) "Dix années de travaux de recherche en évaluation"* (18, 19 et 20 septembre 1996), Grenoble : Université Pierre Mendès France.

- Gilles, J.-L. (1996b). Profils d'auto-évaluation et d'états de connaissance partielle chez les étudiants de première candidature (1995-1996) de la Faculté de Psychologie et des Sciences de l'Éducation de l'Université de Liège, communication présentée lors du *Colloque de l'Association Internationale de Pédagogie Universitaire (A.I.P.U.)*. Tunis : Université de Tunis.
- Gilles, J.-L. (1997). Impact de deux entraînements à l'utilisation des degrés de certitude chez les étudiants de 1ère candidature de la Faculté de Psychologie et des Sciences de l'Éducation de l'ULg, in Boxus, E., Gilles, J.-L., Jans, V. & Leclercq, D. (Eds), *Actes du 15ème 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. (1998a). Apports des mesures métacognitives lors d'un test de compréhension d'un article scientifique, in Depover & Noël (Eds), *Approches plurielles de l'évaluation des processus cognitifs*. Mons : Université de Mons-Hainaut, pp. 19-30.
- Gilles, J.-L. (1998b). Mise en œuvre de tests formatifs à l'aide de l'Internet, in Depover, C. & Noël, B. (Eds), *Approches plurielles de l'évaluation des processus cognitifs*. Mons : Université de Mons-Hainaut, pp. 193-204.
- Gilles, J.-L. & 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., Dabo, M. (1999). Testing Interactif Multimédia via Internet (TIMI) version 1.0 [Computer Software]. Liège : Université de Liège, Faculté de Psychologie et des Sciences de l'Éducation, Système Méthodologique d'Aide à la Réalisation de Tests (SMART).
- Gilles, J.-L., Pirson, M. (1999). Université de Liège - Evaluation des enseignements : données publiques (campagne 1997-1998). Liège : Université de Liège, Commission interfacultaire d'évaluation des enseignements.
- Gilles, J.-L. & 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., 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., Collet, M., Debry, M., Denis, B., Etienne, A.-M., Geuzaine, C., Jans, V., Leclercq, D., Lejeune, M. & Paheau, C. (1998). *Evaluation des enseignements en 1<sup>ère</sup> et 2<sup>ème</sup> candidatures, année académique 1997-1998 - Rapport de synthèse pour le Conseil de Faculté du 10 novembre 1998*. Liège : Université de Liège, Faculté de Psychologie et des Sciences de l'Éducation.
- Gilles, J.-L., Poncin, P., Ruwet, J.-C. et Leclercq, D. (1999). Les travaux dirigés virtuels d'Anthropologie biologique – Bilan d'une première utilisation, in J.-P. Béchard et D. Gregoire (Eds), *Apprendre et enseigner autrement*, Montréal : Ecole des Hautes Études Commerciales, Vol. 1, pp. 294-307.
- Gilles, J.-L., Pirson, M. (2002). Feedbacks Via Internet (FVI) [Computer Software]. Liège : Université de Liège, Faculté de Psychologie et des Sciences de l'Éducation, Système Méthodologique d'Aide à la Réalisation de Tests (SMART).
- Gilles, J.-L. (2002). Qualité spectrale des tests standardisés universitaires – Mise au point d'indices éducatifs d'analyse de la qualité spectrale des évaluations des acquis des étudiants universitaires et application aux épreuves MOHICAN check up '99, thèse de doctorat en Sciences de l'éducation. Liège : Université de Liège, Faculté de psychologie et des sciences de l'éducation de l'université de Liège.

- 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, pp.105-109.
- Gilles, J.-L. (2004). Gérer la qualité des évaluations des acquis des grands groupes d'étudiants à l'aide d'un Système Méthodologique d'Aide à la Réalisation de Tests (SMART), in Arzola, S., Vizcarra, R., Cornejo, A., Undurraga, G. (Eds), *World Education for Educational Research (WAER) - XIV Congreso Mundial de Ciencias de la Educación : Educadores para una Nueva Cultura - Summary of papers abstracts*. Santiago: Pontificia Universidad Católica de Chile, Facultad de Educación.
- Gilles, J.-L. (2004). Analyse d'un dispositif innovant d'ingénierie docimologique permettant aux étudiants d'obtenir des feedbacks individualisés et diagnostiques via le web après une épreuve standardisée en amphithéâtre, in Arzola, S., Vizcarra, R., Cornejo, A., Undurraga, G. (Eds), *World Education for Educational Research (WAER) - XIV Congreso Mundial de Ciencias de la Educación : Educadores para una Nueva Cultura - Summary of papers abstracts*. Santiago: Pontificia Universidad Católica de Chile, Facultad de Educación.
- Gilles, J.-L. et Lovinfosse, V. (2004). Utilisation du cycle SMART de gestion qualité des évaluations standardisées dans le contexte d'une Haute Ecole : regard critique en termes de validité, fidélité, sensibilité des mesures, diagnosticité, praticabilité, équité, communicabilité et authenticité, in Arzola, S., Vizcarra, R., Cornejo, A., Undurraga, G. (Eds), *World Education for Educational Research (WAER) - XIV Congreso Mundial de Ciencias de la Educación : Educadores para una Nueva Cultura - Summary of papers abstracts*. Santiago : Pontificia Universidad Católica de Chile, Facultad de Educación.
- Gilles, J.-L. Piette, S.-A., Detroz, P. et Pirson, M. (2004). Le projet de plate-forme électronique de construction et de gestion qualité de tests standardisés *e-C&QCST*, in Arzola, S., Vizcarra, R., Cornejo, A., Undurraga, G. (Eds), *World Education for Educational Research (WAER) - XIV Congreso Mundial de Ciencias de la Educación : Educadores para una Nueva Cultura - Summary of papers abstracts*. Santiago : Pontificia Universidad Católica de Chile, Facultad de Educación.
- Gilles, J.-L., Lê, H., Pirson, M. (2004). Testing Interactif Multimédia via Internet (TIMI) version 2.0 [Computer Software]. Liège : Université de Liège, Faculté de Psychologie et des Sciences de l'Éducation, Système Méthodologique d'Aide à la Réalisation de Tests (SMART).
- Glasser, W. (1992). *The Quality School*. New York : Harper and Row.
- Guliksen, H. (1950). *Theory of mental tests*. New York : Wiley.
- Hardy, J.-L. (1981). Adjustment of choice tests correlation in multiple choice item analysis, *Série LPE*, 81-07-001.
- Hardy, J.-L. (1983). Plusieurs solutions au problème du recouvrement dans la corrélation entre un test et un item dichotomique, *Scientia Paedagogica Experimentalis*, 49-71.
- Hartog, P. & Rhodes, E.C. (1936). *An Examination of Examinations*. London : Mac Millan.
- Henmon, V. (1911). The relation of the time of a judgment to its accuracy, *Psychological Review*, 18, 186-201.
- Hevner, K.-A. (1932). A method of correcting for guessing in true-false tests and empirical evidence in support of it, *Journal of Social Psychology*, 3, 359-362.
- Hollingworth, R.-L. (1913). Experimental Studies in judgment, *Archives of Psychology*, 29, 1-119.
- Hotelling, H. & Pabst, M.R. (1936). Rank correlation and tests significance involving no assumption of normality. *Ann. Math. Statist.*, 7, 29-43.
- Howell, D.C., (1998). *Méthodes statistiques en sciences humaines*. Bruxelles : De Boeck Université.

- Hunt, D. (1977). *The human self assessment process. Study II : The effects of the number of self-assessment categories on acquisition*. Interim Report from U.S. Army Research Institute for the Behavioral and Social Sciences Grant #DAHC19-76-G-002, New Mexico State University, Las Cruces, NM.
- Jacobs S.-S. (1968). *An empirical investigation of the relationship between selected aspects of personality and confidence-weighting behaviors*, Doctoral dissertation, University of Maryland, University Micro-films, 68, 16, 676.
- Jacobs S.-S. (1971). Correlates of unwarranted confidence in response to objective test items, *Journal of Educational Measurement*, 8, 1.
- 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.
- Jans (2000). *Confrontations Instrumentées et Dialectiques des Jugements Auto- et Alloévaluatifs*, thèse de doctorat en Sciences de l'Education. Liège : Université de Liège, Faculté de Psychologie et des Sciences de l'Education.
- Kane, M. (1994). Validating the performance standards associated with passing scores. *Review of Educational Research*, 64, 425-461.
- Karraker R.J. (1967). Knowledge of results and incorrect recall of plausible multiple choice alternatives, *Journal of Educational Psychology*, 58, 11-14.
- Kauffman, Ch., Dupont, P., Philippart, A. (1995). *Projet pilote européen pour l'évaluation de la qualité dans l'enseignement supérieur*. Bruxelles : Ministère de l'Education, de la Recherche et de la Formation.
- Kaufman, R. (1992). The challenge of total quality management in education. *International Journal of Educational Reform*, 1(2), 149-165.
- Kehoe, J. (1995). *Basic item Analysis for Multiple-Choice Tests*. ERIC/AE Digest. Washington : ERIC Clearinghouse on Assessment and Evaluation, The Catholic University of America, Department of Education, O'Boyle Hall.
- Kelley, T.L. (1939). Selection of upper and lower groups for the validation of test items. *Journal of Educational Psychology*, 30, 17-24.
- Kurst, A.K. & Mayo, S.T. (1979). *Statistical methods in education and psychology*. New York : Springer-Verlag.
- Landercy, A. (1981). *Elements de statistiques*. Mons : Université de Mons, Département d'Édition des Cours.
- Landercy, A. (1983). *Initiation statistique*. Bruxelles : Editions De Boeck.
- Laveault, D. & Gregoire, J. (1997). *Introduction aux théories des tests en sciences humaines*. Bruxelles : De Boeck.
- Laveault, D. (1999). *Autoévaluation et régulation des apprentissages*. in C. Depover & B. Noel (Ed.), *Evaluation des compétences et des processus cognitifs*. Bruxelles : De Boeck, pp. 57-79.
- Leclercq, D. (1973). Critiques de méthodes d'application, de correction et de cotation des questions à choix multiple. *Scientia Paedagogica Experimentalis*, X, 1, 46-57.
- Leclercq, D. (1975). *L'évaluation subjective de la probabilité d'exactitude des réponses en situation pédagogique*. Thèse de doctorat en Sciences de l'Education. Liège : Université de Liège, Institut de Psychologie et des Sciences de l'Education.
- Leclercq, D. (1982). Confidence marking, its use in testing, in B. Choppin & N. Postlethwaite (eds.), *Evaluation in Education : International Review Series*, Oxford : Pergamon, vol. 6, n°2, pp. 161-287.
- Leclercq, D. (1986). *La conception des questions à choix multiple*. Bruxelles : Labor.

- Leclercq, D. (1987). *Qualité des questions et signification des scores avec application aux QCM*. Bruxelles : Labor.
- Leclercq, D. (1993). Validity, Reliability and Acuity of Self-Assessment in Educational Testing, in Leclercq, D. et Bruno, J. (Eds), *Item Banking : Interactive Testing and Self-Assessment*, NATO ASI Series. Heidelberg : Springer Verlag, pp. 113-131.
- Leclercq, D. (1998). *Approche Technologique de l'Education et de la Formation*. Liège : Université de Liège, Service de Technologie de l'Education, 4<sup>ème</sup> édition.
- Leclercq, D. & Detroz, P. (2001). Les résultats globaux du projet MOHICAN, in Leclercq, D. (Ed.), *Le premier des MOHICANs – Une recherche-action de Monitoring Historique des CANDidatures*, Rapport relatif au contrat 798363 entre le Conseil Inter Universitaire Francophone (CIUF) et l'Université de Liège. Bruxelles : CIUF (à paraître).
- Leclercq, D. & Gilles, J.-L. (1993). Hypermedia : Teaching Through Assessment. In D. Leclercq et J. Bruno (Eds), NATO ASI Series, *Item Banking: Interactive Testing and Self Assessment*. Berlin: Springer Verlag, Vol. 112, pp. 31-47.
- Leclercq, D. & Gilles, J.-L. (1994). GUESS, un logiciel pour entraîner à l'auto-estimation de sa compétence cognitive, in A. Dumont et J. Weber (Eds), *Actes du 3ème colloque international ESIEE - Marne-La-Vallée « QCM et questionnaires fermés »*. Paris : Université Paris 7 – Denis Diderot, Laboratoire d'ingénierie didactique, pp.137-158.
- Leclercq, D. & Gilles J.-L. (1995). Le kaléidoscope des techniques de questionnement, *Colloque National de l'Association Internationale de Pédagogie Universitaire (A.I.P.U.)*, Colonster-Liège, 22 septembre 1995.
- Leclercq, D. & Gilles, J.-L. (2001). Techniques de mesure dans l'autoévaluation (dix techniques d'auto-estimation de la qualité de ses réponses), in G. Figari & M. Achouche (Eds), *L'activité évaluative réinterrogée – Regards scolaires et socioprofessionnels*. Bruxelles : Editions De Boeck, pp. 134-142.
- Leclercq, D., Conti, C., De Ketele, J.-M., Delhaxhe, M., Dupont, P., Lambert, J.-P., Lambotte, J.-P., Noël, B., Romainville, M. & Wolfs, J.-L. (2001). Le projet MOHICAN-check-up : son origine, ses intentions, ses méthodes, in Leclercq, D. (Ed.), *Le premier des MOHICANs – Une recherche-action de Monitoring Historique des CANDidatures*, Rapport relatif au contrat 798363 entre le Conseil Inter Universitaire Francophone (CIUF) et l'Université de Liège. Bruxelles : CIUF (à paraître).
- Leclercq, D., Denis, B., Boxus, E. & Gilles, J.-L. (1994). DOUBLE CHECK - Un concept d'évaluation interactive permettant de distinguer les capacités de compréhension et d'analyse, in *Colloque itinérant sur la Pédagogie Universitaire entre les Universités du Québec et les Universités de la Communauté française de Belgique*. Liège : Université de Liège – STE.
- Leclercq, D., Denis, B., Jans, V., Poumay, M. & Gilles, J.-L. (1998). L'amphithéâtre électronique – Une application : le LQRT - SAFE. In D. Leclercq (Ed), *Pour une pédagogie universitaire de qualité*. Liège Sprimont : Mardaga, pp. 161-186.
- Leclercq, D., Georges, F., Gilles, J.-L., Reggers, T., and Rommes, O. (1998b). Interactive Multimedia Programmed Biographies (IMPB) : a new method for clinical training, *proceedings of the BITE (Bringing Information Technology fot Education) conference (25-27 March 1998)*. Maastricht : pp. 406-417.
- Leclercq, D., Jans, V., Georges, F. & Gilles, J.-L. (2000). Objective assessment of subjectivity : applying confidence marking to partial knowledge, Communication présentée à la conférence EARLI SIG Assessment, Maastricht, septembre 2000.
- Leclercq, D., Peeters, R., Reggers, T., Charlier, J.E., De Ketele, J.M., Delhaxhe, M., Dupont, P., Lambert, J.P., Lambotte, J.P., Loeckx, E., Pilatte, A., Romainville, M. & Wolfs, J.-L. (1997). *Franchir le cap des candis – Inscriptions et réussites en candidatures universitaires dans la Communauté Française de Belgique de 1988 à 1995*. Bruxelles : Conseil Interuniversitaire de la Communauté Française.

- Leclercq, D., Rommes, O., Georges, F. et Gilles J.-L. (1998). Une préparation multimédias à affronter les conflits professeur-élèves en classe, *Education formation*, pp. 90-93.
- Leclercq, D., Willain, J.-C. , Denis, B., Poumay, M., Gilles, J.-L., Orban, M. and Jans, V. (1999). Votes en amphithéâtre électronique pour animer de grands auditorios universitaires selon six paradigmes d'apprentissage/enseignement, in J.-P. Bécharde et D. Gregoire (Eds), *Apprendre et enseigner autrement*. Montréal : Ecole des Hautes Etudes Commerciales, Vol 2, pp. 567-578.
- Leclercq, D., Dupont, C., Detroz, P. et Gilles J.-L. (2002). Faire mentir les prédictions en 1<sup>ère</sup> candidature : comment ? Avec quel résultat ? L'opération RESSAC (Résultats d'épreuves standardisées au service de l'apprentissage en candis), *Actes du 2<sup>ème</sup> Congrès des chercheurs en éducation, l'école dans quel(s) sens ? (12 et 13 mars 2002)*. Bruxelles : Ministère de l'Enseignement supérieur, de l'Enseignement de Promotion sociale et de la Recherche scientifique, pp. 39-42.
- Leclercq, D., Detroz, P., Dupont, C. et Gilles J.-L. (2003). L'opération RESSAC : diagnostic métacognitif avant la période de "bloque", *Actes du 16<sup>ème</sup> Colloque international de l'ADMEE-Europe (4 au 6 septembre 2003)*. Liège : Département Education et Formation de l'Université de Liège.
- Leclercq, D. et Gilles J.-L. (2003). Analyses psychométriques des questions des 10 check-up MOHICAN - Vue d'ensemble, in Leclercq (Ed), *Diagnostic cognitif et métacognitif au seuil de l'université - Le projet MOHICAN mené par les 9 universités de la Communauté française Wallonie Bruxelles*. Liège : Presses universitaires de l'Université de Liège.
- Leclercq, D., Georges, F., Gilles J.-L., Pirson, M. et Simon, F. (2003). Les feedbacks des check-up MOHICAN aux étudiants et aux enseignants, in Leclercq (Ed), *Diagnostic cognitif et métacognitif au seuil de l'université - Le projet MOHICAN mené par les 9 universités de la Communauté française Wallonie Bruxelles*. Liège : Presses universitaires de l'Université de Liège.
- Leclercq, D. Detroz, P., Dupont, C. et Gilles J.-L. (2003). Changer de méthode d'étude suite aux feedbacks : L'opération RESSAC (Résultats d'Épreuves Standardisées au Service de l'Apprentissage en Candidature), in Leclercq (Ed), *Diagnostic cognitif et métacognitif au seuil de l'université - Le projet MOHICAN mené par les 9 universités de la Communauté française Wallonie Bruxelles*. Liège : Presses universitaires de l'Université de Liège.
- Leclercq, D., Gilles J.-L. et Dupont, C. (2003). L'information des feedbacks pour les professeurs - L'opération VOCA PROFS à l'ULg ou Estimation par les professeurs des résultats des étudiants au check-up de Vocabulaire, in Leclercq (Ed), *Diagnostic cognitif et métacognitif au seuil de l'université - Le projet MOHICAN mené par les 9 universités de la Communauté française Wallonie Bruxelles*. Liège : Presses universitaires de l'Université de Liège.
- Lichtenstein, S., Fischhoff, B. & Philips, L. (1977). Calibration of probabilities : The state of the art, in Jungermann & Dezeeuw (Eds), *Decision making and change in human affairs. Proceedings of the 5<sup>th</sup> SPUDM Conference*. Darmstadt : Reidel Publishing Company, 275-324.
- Lion, Y. (2000). Tests chez les ingénieurs. Communication présentée lors des journées Form@sup, Liège : Université de Liège, Service de Technologie de l'Éducation.
- Lord, F.M. & Novick, M.R. (1968). *Statistical Theories of Mental Test Scores*. Reading, Mass. : Addison-Wesley.
- Mac Guigan, F.J. (1967). The G Statistics, An Index of Amount Learned. *N.S.P. Journal*, 69, pp. 14-16.
- Magnuson (1967). *Test theory*. Boston : Addison-Wesley.
- Messick (1988). The once and the future issues of validity : Assessing the meaning and consequences of measurement. In H. Wainer & H.I. Braun (Eds), *Test Validity*. Hillsdale, NJ : Lawrence Erlbaum.
- Michael, J.-J. (1968). The reliability of a multiple-choice examination under various test-making instructions. *Journal of Educational Measurement*, 5, 307-314.

- Miller, G.-A. (1956). The magical number of seven, plus minus two, *Psychological Review*, Vol 63, 81-97.
- Murphy A.H. & Winkler R.L. (1974). Subjective probability forecasting experiments in meteorology : some preliminary results. *Bulletin of the American Meteorological Society*, 55, 1206-1216, 1974.
- Nightingale, P. & O'Neil, M. (1994). *Achieving quality learning in higher education*. London : Kogan Page.
- Nightingale, P. & O'Neil, M. (1994). *Achieving quality learning in higher education*. London : Kogan Page.
- Nitko, A. (1996). *Educational Assessment of Students*. Englewood Cliffs : Merrill, second edition.
- Oskamp, S. (1962). The relationship of clinical experience and training methods to several criteria of clinical prediction, *Psychological Monographs*, 76.
- Passeron, J.C. (1970). Sociologie des examens. *Education et Gestion*, 2, 6-16.
- Phillips, C. (1996). "Towards a Task-Based Methodology for Designing GUIs." In Purvis, M, ed., *Software engineering: Education & practice*. Los Alamitos, CA:IEEE Computer Society Press.
- Piaget, J. (1969). *Psychologie et pédagogie*. Paris : Denoël.
- Pieron, H., Reuchlin, M. & Bacher, F. (1962). Une recherche expérimentale de docimologie sur des examens oraux de physique au niveau du baccalauréat de mathématiques, *Biotypologie*, 23, 48-73.
- Pieron, H. (1963). *Examens et docimologie*. Paris, Presses Universitaires de France.
- Pitz, G.F. (1974). Subjective probability distributions for imperfectly known quantities. In Gregg, L.W. (Ed.) *Knowledge and Cognition*. New York : Wiley, pp. 29-41.
- Preston, R.C. (1965). Multiple-choice test as an instrument in perpetuating false concepts, *Educational and Psychological Measurement*, 34, 499-509.
- Ramsden, P. (1991). A performance indicator of teaching quality in higher education : the course experience questionnaire, *Studies in Higher Education*, 16, 129-150.
- Rasch, G. (1966). An item analysis which takes individual differences into account, *British Journal of Mathematical and Statistical Psychology*, 19(1), 49-57.
- Rhodes, L.A. (1992). *On the road to quality*. Educational leadership, 49(6), 76-80.
- Romainville, M. & Boxus, E. (1998). La qualité en pédagogie universitaire. In D. Leclercq (Ed), *Pour une pédagogie universitaire de qualité*. Liège Sprimont : Mardaga, pp. 13-32.
- Sandbergen (1971). Guessing and Confidence in testing educational achievement, In Choppin, B., *A/106 IEA Memorandum*.
- Segers, M. & Dochy, F. (1996). Quality assurance in higher education : theoretical considerations and empirical evidences, *Studies in Educational Evaluation*, 1996, 22 (2), 115-137.
- Shufford, A. (1993). In Pursuit of the Fallacy : Resurrecting the Penalty, in D. Leclercq & J. Bruno (Eds), *Item Banking : Interactive Testing and Self-Assessment*, NATO ASI Series, Berlin : Springer Verlag, pp. 76-98.
- Shufford, A., Albert, A. & Massengil, N.E. (1966). Admissible probability measurement procedures, *Psychometrika*, 31, 125-145.
- Skinner, B.F. (1961). Why we need teaching machines, *Harvard Educational Review*, 31, 377-398.
- Spearman, C. (1907). Demonstration of formulae for true measurement of correlation. *American Journal of Psychology*, 18, 161-169.
- Swineford, F. (1938). The measurement of a personality trait., *Journ. of Educ. Psych.*, 29, 289-292.

- Tourneur, Y. (1988). *La psychométrie et la construction des instruments d'évaluation - Recueil de textes*. Mons : Université de Mons, Département d'Édition des Cours.
- Trow W.-C. (1923). The psychology of confidence, an experimental inquiry, *Archives of Psychology*, 67, 1-47.
- Van Der Vleuten, C. & Wijnen, W. (1990). *Problem-Based learning : Perspective from the Maastricht experience*. Amsterdam : Thesis.
- Van Der Vleuten, C. (1996). *Beyond intuition, inaugural lecture*. Maastricht : Universitaire Press Maastricht.
- Van Lenthe, J. (1993). The Development and Evaluation of ELI, an Interactive Elicitation Technique for Subjective Probability Distributions, in Leclercq, D. et Bruno, J. (Eds), *Item Banking : Interactive Testing and Self-Assessment*, NATO ASI Series. Heidelberg : Springer Verlag.
- Van Naerssen, R.-F. & Van Beaumont, E. (1965). Ervaringen met een Zekerheidsaanduiding bij objectieve Tentamen, *Nederlands Tijdschrift Psychologie*, 20, 308-315.
- Van Naerssen, R.F. (1962). A scale for measurement of subjective probability, *Acta Psychologica*, 20, 2, 159-166.
- Vandevelde, L. (1971). *Réflexions sur les QCM*. Bruxelles : ULB, document ronéotyp.
- Wiersma, W. & Jurs, S.G. (1990). *Educational Measurement and Testing*. Boston : Allyn and Bacon.
- Wiley, L.-N. & Trimble, D.-C. (1936). The ordinary objective test as a possible criterion of certain personality, *School and Society*, 43, 446-448.
- Wirfs-Brock, R. (1993). "Designing Scenarios: Making the Case for a Use Case Framework" *Smalltalk Report*, November-December.
- Wood, R. (1977). Multiple-choice: A state of the art report, in Choppin and Postlethwaite (Eds), *Evaluation in Education International Progress*. Oxford : Pergamon.
- Wright, B. & Stone (1979). *Best test Design*, Chicago IL : MESA Press.
- Zink, K.J. & Schmidt, A. (1995). Measuring universities against the european quality award criteria. *Total quality management*, 6 (5-6), 547-561.

## ANNEXES

**Annexe I :** Dias PowerPoint de la présentation « electronic Construction and Quality Control in Standardized Testing, the e-C&QCST project » effectuée à la 11<sup>ème</sup> conférence biennale de EARLI



EARLI 11th Biennial conference 2005



## electronic Construction and Quality Control in Standardized Testing: the e-C&QCST project

*Jean-Luc Gilles, Sylvie-Anne Piette, Pascal Detroz, Salvatore Tinnirello,  
Marc Pirson, Mbassa Dabo & Hung Lê*

<http://www.smart.ulg.ac.be/earli-2005/e-CQCST.ppt>



University of Liège - Belgium



## I. Context

- Many European Countries have to face a growing number of higher education students
- The training professionals and the human resources managers of companies are often constrained to evaluate staff competencies, often acquired outside the schools
- There is a worldwide movement of the introduction of quality management in education and training activities



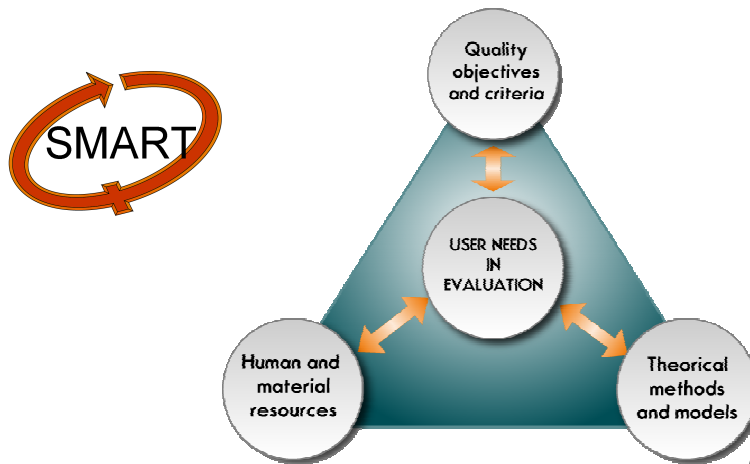
*(Kaufman & al., 1995; Segers & Dochy, 1996; Ramsden, 1991; Nightingale & O'Neil, 1994; Zink & Schmidt, 1995; Castaigne, Gilles & Hansen, 2001; Gilles, 2002; Gilles et Lovinfosse, 2004; Phelps, 2005)*

EARLI 11th Biennial conference 2005 – electronic Construction and Quality Control in Standardized Testing: the e-C&QCST project  
 Jean-Luc Gilles, Sylvie-Anne Piette, Pascal Detroz, Salvatore Tinnirello, Marc Pirson, Mbassa Dabo & Hung Lê



## I. Context

- The quality approach suggested by the SMART team in this context



*(Gilles, 2002)*

EARLI 11th Biennial conference 2005 – electronic Construction and Quality Control in Standardized Testing: the e-C&QCST project  
 Jean-Luc Gilles, Sylvie-Anne Piette, Pascal Detroz, Salvatore Tinnirello, Marc Pirson, Mbassa Dabo & Hung Lê





## II. Principals R&D objectives of the e-C&QCST project:

- To develop a web based platform which propose methods, procedures and tools for the management of standardized tests.
- To set up economic valorizations of the results of our research with supply of services like:
  - Audit of standardized testing systems
  - Training and coaching of the users of the e-C&QCST platform
  - Docimological services for the improvement of standardized tests
  - Test creation with the platform
- To propose a quality label for standardized testing

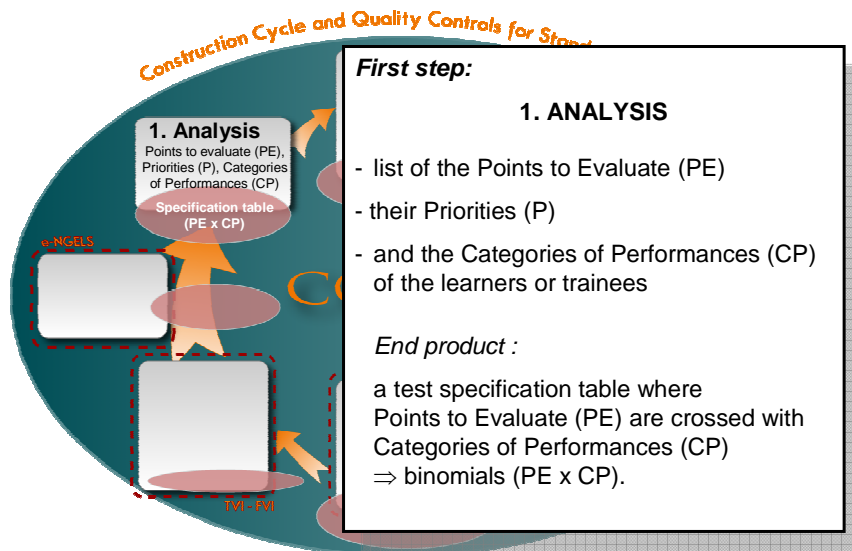
EARLI 11th Biennial conference 2005 – electronic Construction and Quality Control in Standardized Testing: the e-C&QCST project  
 Jean-Luc Gilles, Sylvie-Anne Piette, Pascal Detroz, Salvatore Tinnirello, Marc Pirson, Mbassa Dabo & Hung Lê



4



## III. Method: the CCQCST model

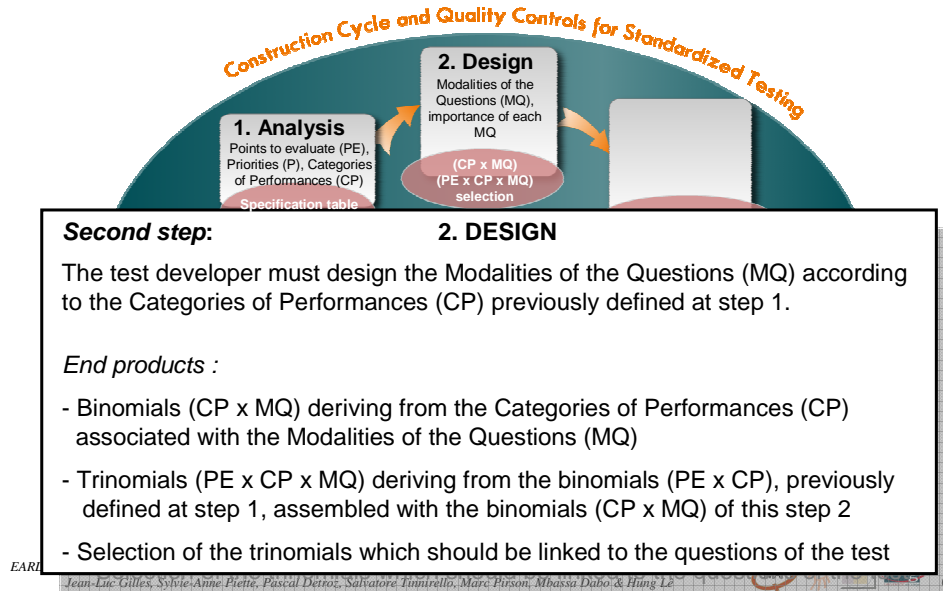


EARLI 11th Biennial conference 2005 – electronic Construction and Quality Control in Standardized Testing: the e-C&QCST project  
 Jean-Luc Gilles, Sylvie-Anne Piette, Pascal Detroz, Salvatore Tinnirello, Marc Pirson, Mbassa Dabo & Hung Lê



5

## III. Method: the CCQCST model



**1. Analysis**  
 Points to evaluate (PE),  
 Priorities (P), Categories  
 of Performances (CP)  
 Specification table

**2. Design**  
 Modalities of the  
 Questions (MQ),  
 importance of each  
 MQ  
 (CP x MQ)  
 (PE x CP x MQ)  
 selection

**Second step: 2. DESIGN**

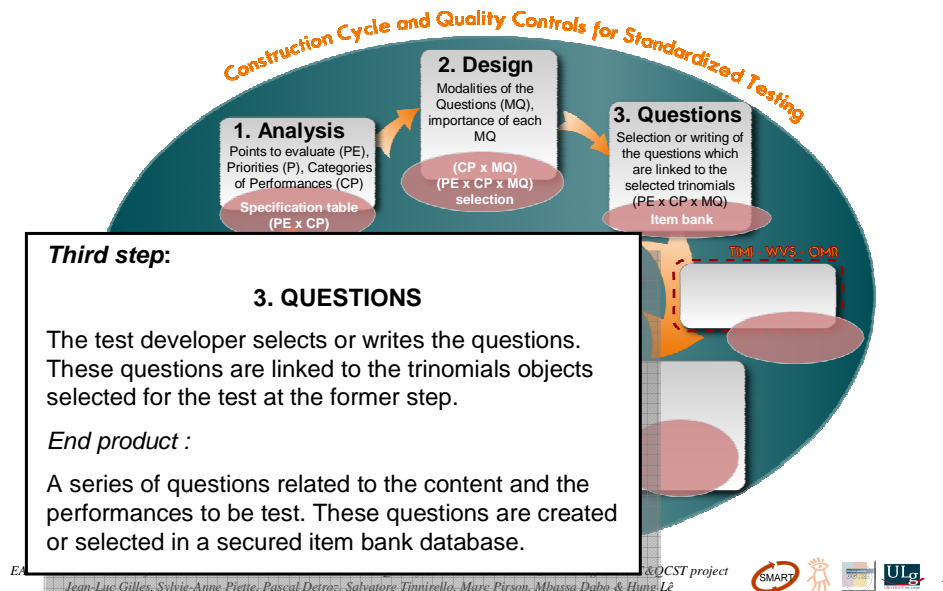
The test developer must design the Modalities of the Questions (MQ) according to the Categories of Performances (CP) previously defined at step 1.

*End products :*

- Binomials (CP x MQ) deriving from the Categories of Performances (CP) associated with the Modalities of the Questions (MQ)
- Trinomials (PE x CP x MQ) deriving from the binomials (PE x CP), previously defined at step 1, assembled with the binomials (CP x MQ) of this step 2
- Selection of the trinomials which should be linked to the questions of the test

EARI Jean-Luc Gilles, Sylvie-Anne Piette, Pascal Detroz, Salvatore Tinnirello, Marc Pirson, Mbassa Dabo & Hung Lê

## III. Method: the CCQCST model



**1. Analysis**  
 Points to evaluate (PE),  
 Priorities (P), Categories  
 of Performances (CP)  
 Specification table  
 (PE x CP)

**2. Design**  
 Modalities of the  
 Questions (MQ),  
 importance of each  
 MQ  
 (CP x MQ)  
 (PE x CP x MQ)  
 selection

**3. Questions**  
 Selection or writing of  
 the questions which  
 are linked to the  
 selected trinomials  
 (PE x CP x MQ)  
 Item bank

**Third step: 3. QUESTIONS**

The test developer selects or writes the questions. These questions are linked to the trinomial objects selected for the test at the former step.

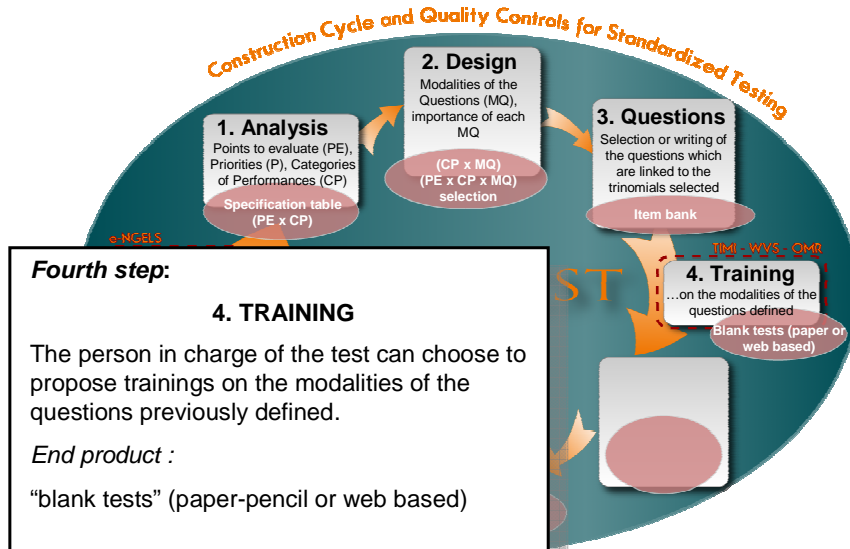
*End product :*

A series of questions related to the content and the performances to be test. These questions are created or selected in a secured item bank database.

EARI Jean-Luc Gilles, Sylvie-Anne Piette, Pascal Detroz, Salvatore Tinnirello, Marc Pirson, Mbassa Dabo & Hung Lê



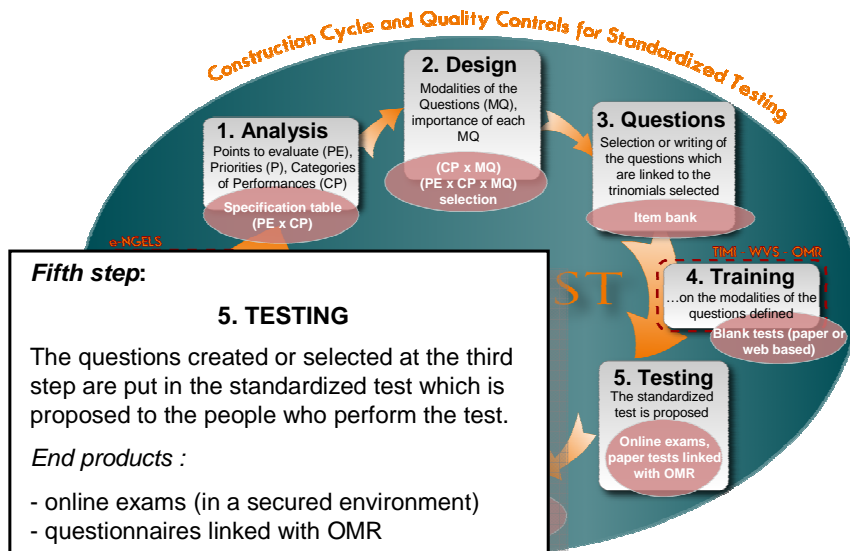
## III. Method: the CCQCST model



EARLI 11th Biennial conference 2005 - electronic Construction and Quality Control in Standardized Testing: the e-C&QCST project  
 Jean-Luc Gilles, Sylvie-Anne Piette, Pascal Detroz, Salvatore Tinnirello, Marc Pirson, Mbassa Dabo & Hung Lê



## III. Method: the CCQCST model

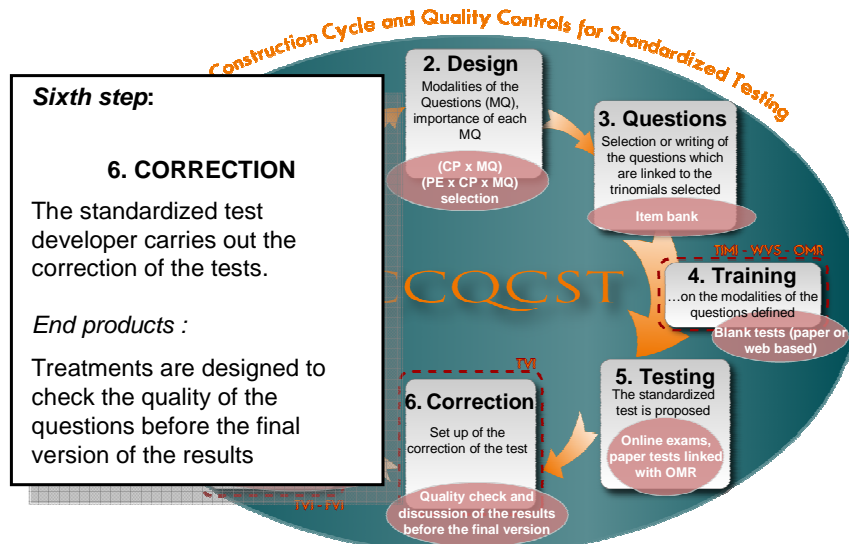


EARLI 11th Biennial conference 2005 - electronic Construction and Quality Control in Standardized Testing: the e-C&QCST project  
 Jean-Luc Gilles, Sylvie-Anne Piette, Pascal Detroz, Salvatore Tinnirello, Marc Pirson, Mbassa Dabo & Hung Lê





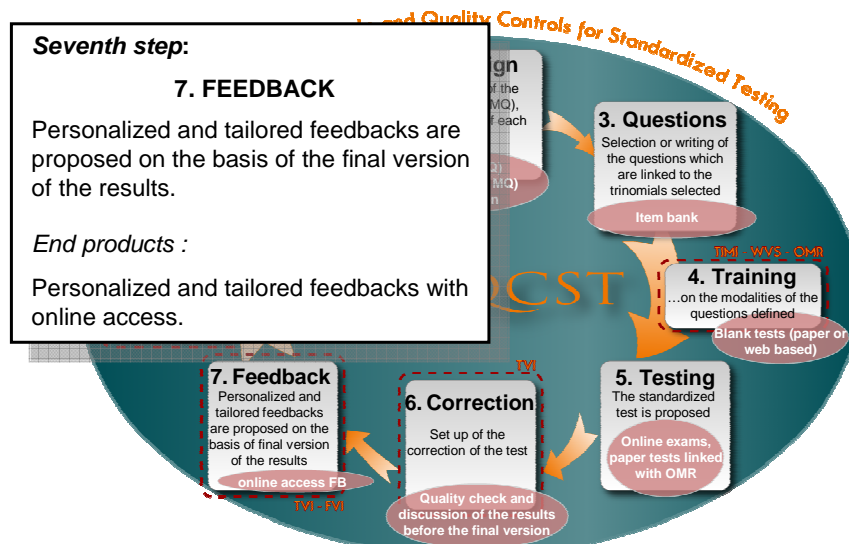
## III. Method: the CCQCST model



EARLI 11th Biennial conference 2005 – electronic Construction and Quality Control in Standardized Testing: the e-C&QCST project  
 Jean-Luc Gilles, Sylvie-Anne Piette, Pascal Detroz, Salvatore Tinnirello, Marc Pirson, Mbassa Dabo & Hung Lê



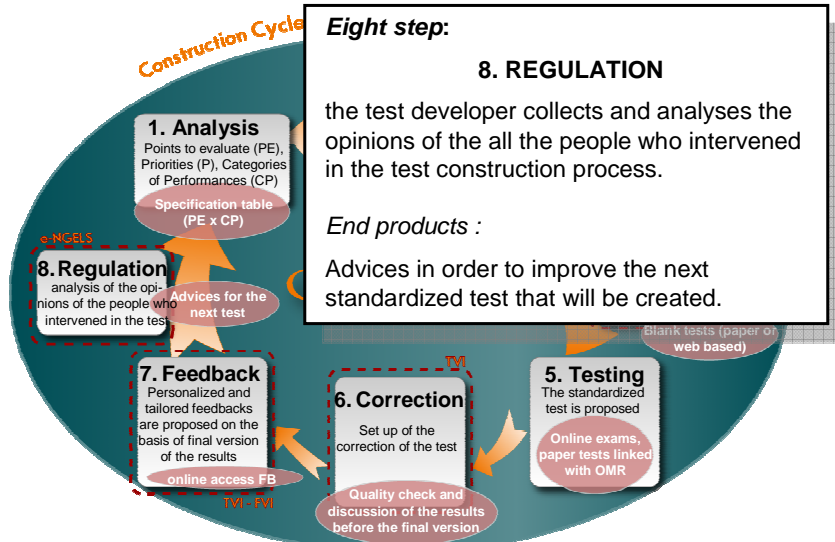
## III. Method: the CCQCST model



EARLI 11th Biennial conference 2005 – electronic Construction and Quality Control in Standardized Testing: the e-C&QCST project  
 Jean-Luc Gilles, Sylvie-Anne Piette, Pascal Detroz, Salvatore Tinnirello, Marc Pirson, Mbassa Dabo & Hung Lê

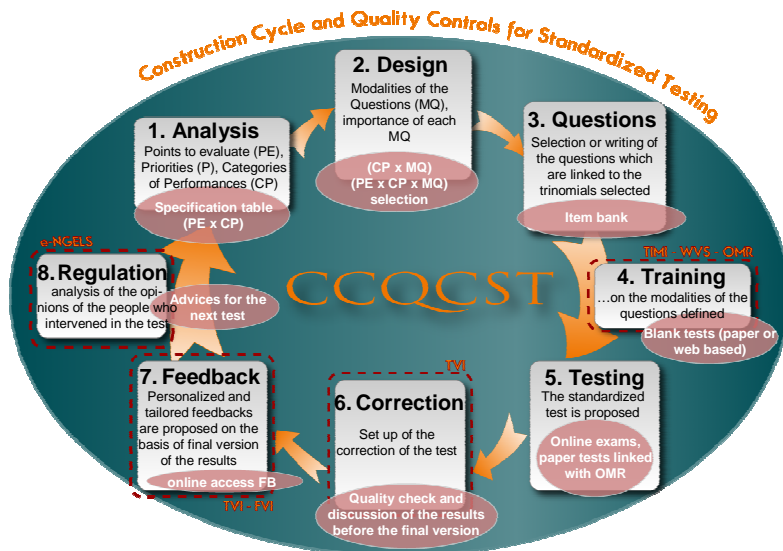


III. Method: the CCQCST model



EARLI 11th Biennial conference 2005 – electronic Construction and Quality Control in Standardized Testing: the e-C&QCST project  
Jean-Luc Gilles, Sylvie-Anne Piette, Pascal Detroz, Salvatore Tinnirello, Marc Pirson, Mbassa Dabo & Hung Lê

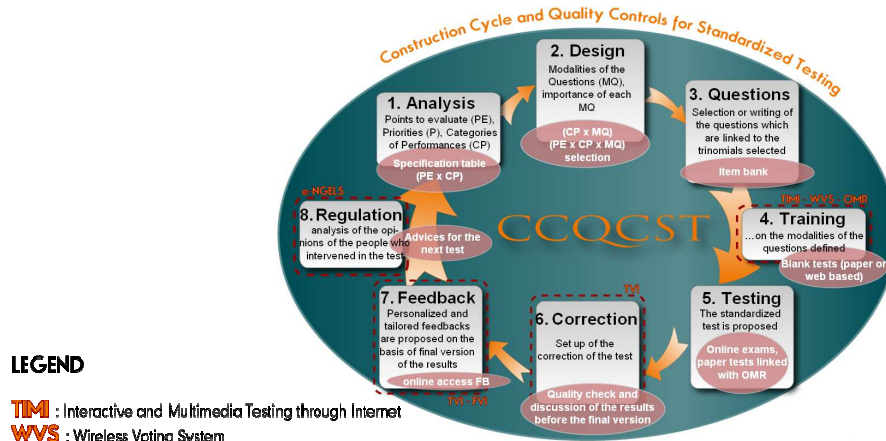
III. Method: the CCQCST model



EARLI 11th Biennial conference 2005 – electronic Construction and Quality Control in Standardized Testing: the e-C&QCST project  
Jean-Luc Gilles, Sylvie-Anne Piette, Pascal Detroz, Salvatore Tinnirello, Marc Pirson, Mbassa Dabo & Hung Lê

e-C&QCST  
 ELECTRONIC CONSTRUCTION AND QUALITY CONTROL IN STANDARDIZED TESTING

## III. Method: the CCQCST model



EARLI 11th Biennial conference 2005 – electronic Construction and Quality Control in Standardized Testing: the e-C&QCST project  
 Jean-Luc Gilles, Sylvie-Anne Piette, Pascal Detroz, Salvatore Tinnirello, Marc Pirson, Mbassa Dabo & Hung Lê



e-C&QCST  
 ELECTRONIC CONSTRUCTION AND QUALITY CONTROL IN STANDARDIZED TESTING

## V. Companies & organizations collaborating in this project



M. Bernard Van Parys  
 M. Hubert de Broqueville



M. Alain Kock



M. Francis Lajoie  
 M. Jean-Pierre Renault



Mme Régine Gendarme  
 Mme Catherine Deruisseau  
 Mme Rameault



M. Dominique Pierret  
 M. Jean-Marie Pairoux



Mme Sandra Schillemans  
 Mme Tania Delhaise  
 M. Pierre Ramelot



Mme Maryse Riffont  
 M. Peter Vanderheyden



Mme Catherine Warnant  
 M. Jean-Philippe Léonard



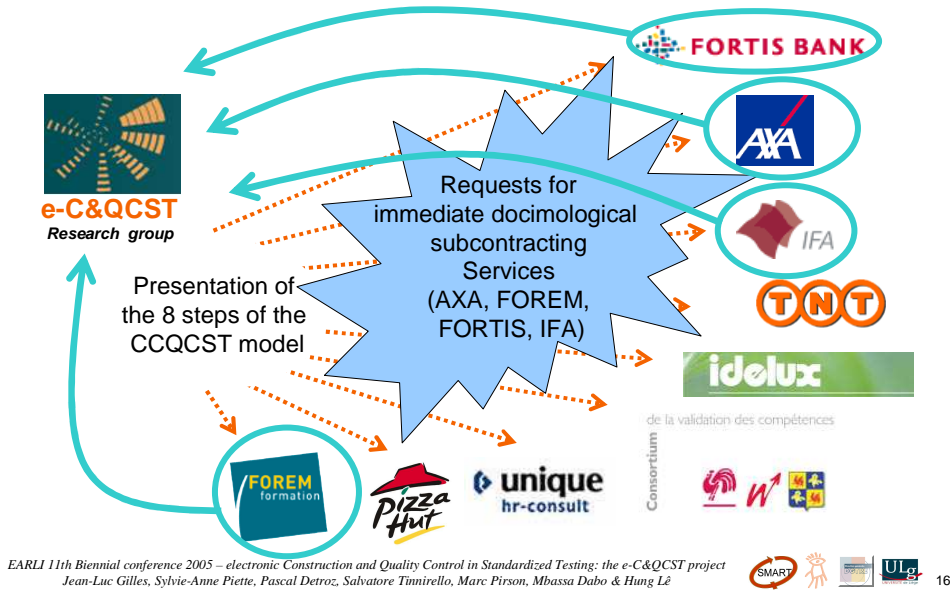
M. Yvan Lagage

EARLI 11th Biennial conference 2005 – electronic Construction and Quality Control in Standardized Testing: the e-C&QCST project  
 Jean-Luc Gilles, Sylvie-Anne Piette, Pascal Detroz, Salvatore Tinnirello, Marc Pirson, Mbassa Dabo & Hung Lê





VI. First results: Relevance of the CCQCST model



VI. First results: Relevance of the CCQCST model

- **Interviews of tests developers of the ten companies and organizations:**
  - Description of the actions;
  - Collaborators concerned;
  - Temporal dimension of the tests construction process;
  - Other processes in which tests are inserted;
  - Importance of the quality of the standardized tests.
  - ...
- **First results of the interviews and the immediate subcontracting services:**
  - The CCQCST model and its procedures are considered as relevant and well accepted;
  - The principal readjustment we have to plan consists of an integration of different roles of tests developers working in teams.



## VI. First results: Users' roles within e-C&QCST platform

- Two difficulties for the users' roles definition:
  - They must be used in very different companies' or non academic organizations' contexts;
  - They should be assigned by one person in charge of the test (the head of the project) to a series of collaborators or subcontractors.
- We propose seven roles within the e-C&QCST platform:
  - Role 1: "INSTALLER" ⇒ *installs the software on the server, ... ;*
  - Role 2: "ADMINISTRATOR" ⇒ *defines the different tests supervisors, ... ;*
  - Role 3: "TEST SUPERVISOR" ⇒ *defines and supervises his collaborators & subcontractors for the tasks related to the test construction, ... ;*
  - Role 4: "TEST DEVELOPER" ⇒ *realizes one or more tasks of one or more steps of the CCQCST model, ... ;*
  - Role 5: "QUALITY SUPERVISOR" ⇒ *defines and supervises his collaborators or subcontractors in charge of the quality controls, ... ;*
  - Role 6: "QUALITY CONTROLLER" ⇒ *realizes one or more tasks related to the quality controls of the productions of the tests developers, ... ;*
  - Role 7: "ASSEESSEE" ⇒ *performs the standardized test.*

EARLI 11th Biennial conference 2005 – electronic Construction and Quality Control in Standardized Testing: the e-C&QCST project  
 Jean-Luc Gilles, Sylvie-Anne Piette, Pascal Detroz, Salvatore Tinnirello, Marc Pirson, Mbassa Dabo & Hung Lê



18



## VI. First results: The SAFES project

- Being given the success of the subcontracting services requested by four companies, we filed in a FIRST SPIN-OFF project of the DGTRÉ.
- Our project entitled "Smart Advices For Evaluation Systems (SAFES)" was well ranked by an independent scientific committee so that we are currently, starting a two years feasibility study that will allows us :
  - To create a method of audit of standardized testing systems;
  - To improve the quality control procedures to be implemented;
  - To define a typology of subcontracted services;
  - To realize a market study.
- The creation of the spin-off company should coincide with the finalisation of the e-C&QCST platform.

EARLI 11th Biennial conference 2005 – electronic Construction and Quality Control in Standardized Testing: the e-C&QCST project  
 Jean-Luc Gilles, Sylvie-Anne Piette, Pascal Detroz, Salvatore Tinnirello, Marc Pirson, Mbassa Dabo & Hung Lê



19



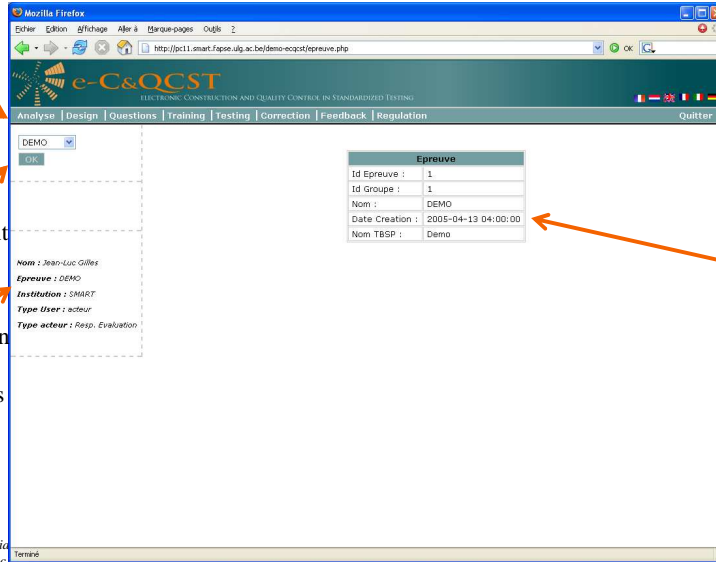
## VI. First results: First interfaces of the e-C&QCST platform

- General ergonomics:

Main menu with the 8 steps of the CCQCST model

Assessment selector

Information about the user who is connected



Multi lingual menu

Information about the assessment which is selected (an assessment includes one or more standardised tests)

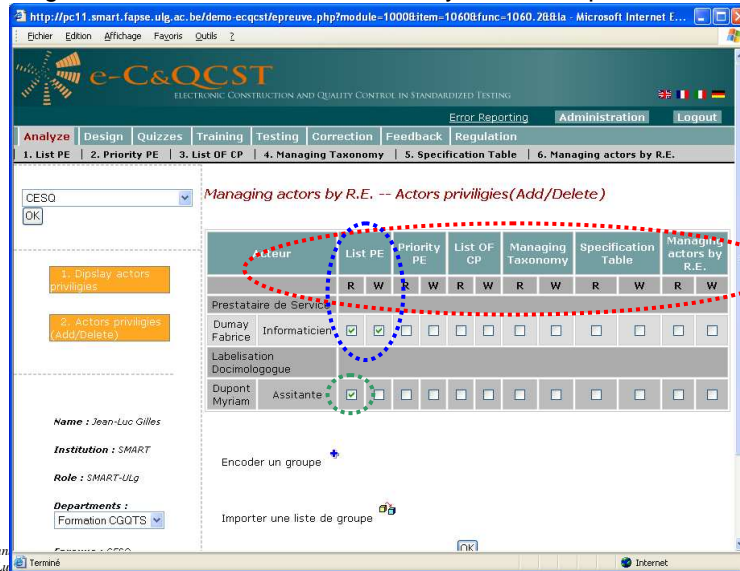
EARLI 11th Bienn  
Jean-Luc

20



## VI. First results: First interfaces of the e-C&QCST platform

- Management of the e-C&QCST users by the test supervisor:



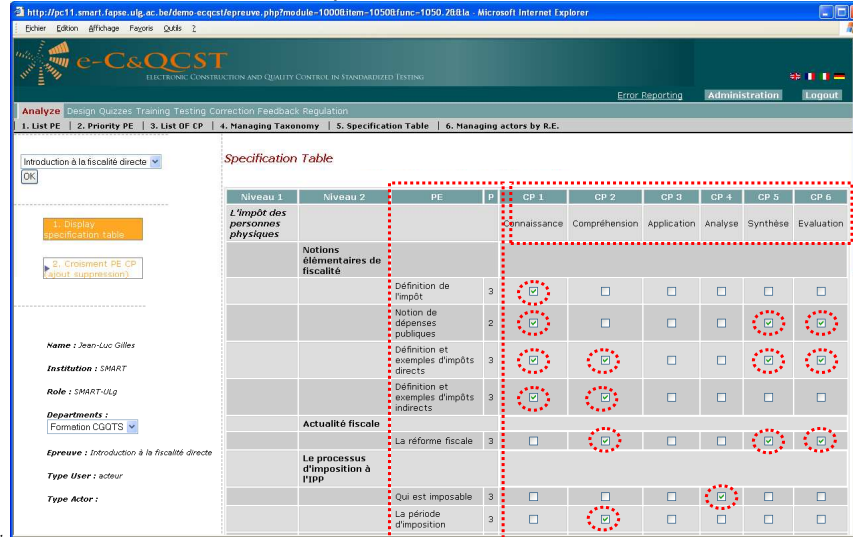
EARLI 11th Bienn  
Jean-Luc

21



## VI. First results: First interfaces of the e-C&QCST platform

- 1st module "ANALYSIS" - specification table: PE are crossed with CP



Niveau 1	Niveau 2	PE	P	CP 1	CP 2	CP 3	CP 4	CP 5	CP 6
L'impôt des personnes physiques				Connaissance	Compréhension	Application	Analyse	Synthèse	Evaluation
	Notions élémentaires de fiscalité								
		Définition de l'impôt	3	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
		Notion de dépenses publiques	2	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
		Définition et exemples d'impôts directs	3	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
		Définition et exemples d'impôts indirects	3	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Actualité fiscale								
		La réforme fiscale	3	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
	Le processus d'imposition à l'IPP								
		Qui est imposable	3	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
		La période d'imposition	3	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

EARLI 11th Jean-Luc Gilles, Sylvie-Anne Piette, Pascal Detroz, Salvatore Tinnirello, Marc Pirson, Mbassa Dabo & Hung Lê



## VII. Conclusions

- Docimological framework:**  
The CCQCST model is relevant and well appreciated in the context of the ten companies and organisations which are our partners in this project
- Ergonomics aspects:**  
The first results show that the ergonomics of the first interfaces facilitate the production of specification tables related to the first step of the standardized tests construction
- Economical valorization framework:**  
The surprising immediate docimological subcontracting services requested by four of the ten companies (also the acceptance of the SAFES project) shows that there are possibilities for economical valorisations of the results of this project.
- Users framework:**  
The seven categories of users' roles we defined were validated by our partners. This clarification should contribute to highlight the complementarities and the responsibilities of the persons who are involved in standardized test creation.

EARLI 11th Biennial conference 2005 - electronic Construction and Quality Control in Standardized Testing: the e-C&QCST project  
 Jean-Luc Gilles, Sylvie-Anne Piette, Pascal Detroz, Salvatore Tinnirello, Marc Pirson, Mbassa Dabo & Hung Lê

Thank you for your attention !

For more information concerning the  
electronic Construction and Quality Control  
in Standardized Tests project  
(e-C&QCST)

do not hesitate to contact me:

jl.gilles@ulg.ac.be

<http://www.smart.ulg.ac.be/earli-2005/e-CQCST.ppt>

**Annexe II :** Dias PowerPoint de la présentation « *Development of edumetrical indices for the analysis of the spectral quality of higher education standardized tests* » effectuée à la 11<sup>ème</sup> conférence biennale de EARLI



*EARLI 11th Biennial conference 2005*



## Development of edumetrical indices for the analysis of the spectral quality of higher education standardized tests

*Jean-Luc Gilles*

<http://www.smart.ulg.ac.be/earli-2005/spectral-analysis.ppt>

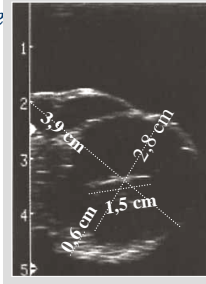


*Système Méthodologique d'Aide à la Réalisation de Tests  
University of Liège - Belgium*



## Introduction:

**Example of Multiple Choice Question:**  
 (Hansen & al, 2000)



This ultrasound of the left ovary of a 6 years old cow, has been realized one week after the oestrus.

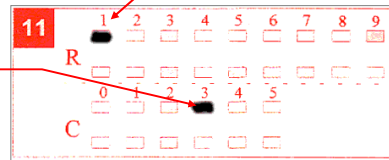
Diagnostic :

1. a follicle of de Graaf
2. a follicular cyst
3. a yellow body
4. a cavitary yellow body
5. an artifact
6. none
7. all

**Example of confidence degrees scale** (Leclercq & Gilles, 2000):  
 The student is invited to accompany his choice by the percentage of chances that his answer will be correct.

If you consider you have a probability of correct answer of	Check
0 %	0
20 %	1
40 %	2
60 %	3
80 %	4
100 %	5

With an optical mark reading device it is possible to read automatically answers and confidence degrees of hundreds of students:



EARLI 11th Biennial conference 2005 - Development of edumetrical indices for the analysis of the spectral quality of higher education standardized tests - Jean-Luc Gilles 2

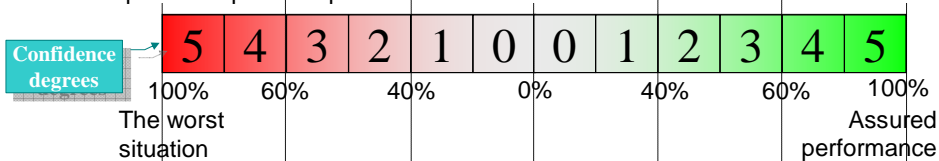
## Introduction:

- Without confidence degrees, the selected solution of a MCQ is either correct or incorrect.

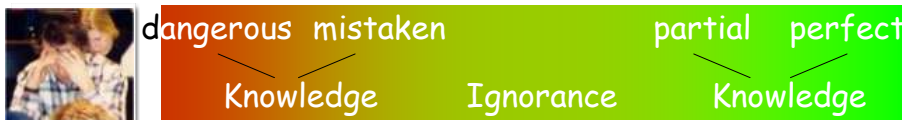
The answer is:

incorrect	correct
-----------	---------

- With the confidence degrees technique it is possible to bypass the "binary" character of students assessment performance and to consider a more spectral conception of a person's performances



- With the confidence degrees technique it is possible to deliver more subtle feedback on each person's spectral performances!



(Jans & Leclercq, 1999)

EARLI 11th Biennial conference 2005 - Development of edumetrical indices for the analysis of the spectral quality of higher education standardized tests - Jean-Luc Gilles 3

**Starting intuition for the new indices used for highlighting problematic MCQs and, within those, the proposals which contain anomalies:**

**Confidence degrees of the students who answer incorrectly**

**Confidence degrees of the students who answer correctly**

We have exploited the confidence degrees provided by the students to provide spectral information on the quality of the questions (as opposed to information on the quality of students' performances).

This ultrasound of the left ovary of a 6 years old cow, has been realized one week after the oestrus.  
Diagnostic :  
1. a follicle of de Graaf  
2. a follicular cyst  
3. a yellow body  
4. a cavitary yellow body  
5. an artifact  
6. none  
7. all

## Spectral quality of the **question**

EARLI 11th Biennial conference 2005 - Development of edumetrical indices for the analysis of the spectral quality of higher education standardized tests - Jean-Luc Gilles 4

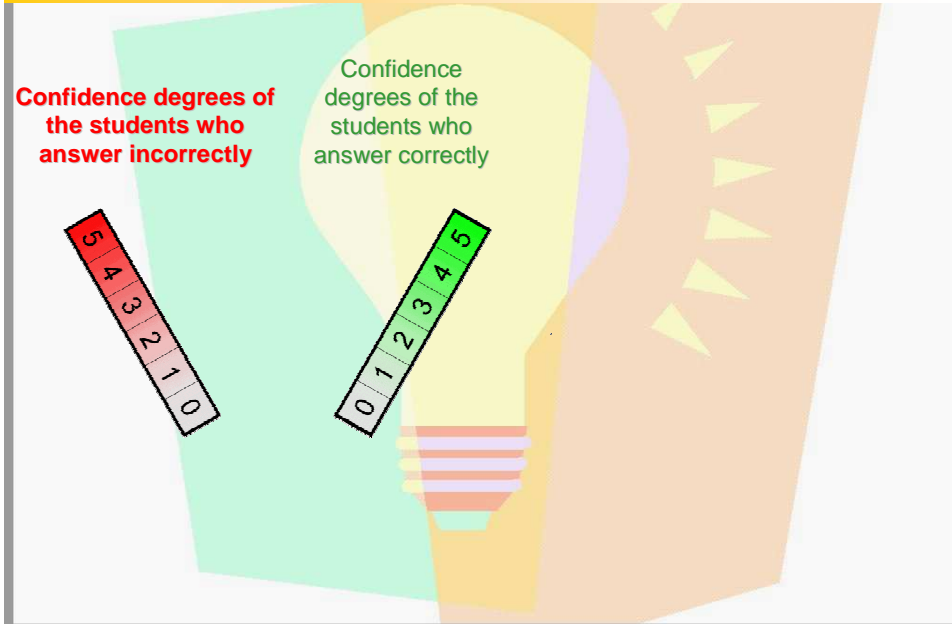
**Starting intuition for the new indices used for highlighting problematic MCQs and, within those, the proposals which contain anomalies:**

**Confidence degrees of the students who answer incorrectly**

**Confidence degrees of the students who answer correctly**

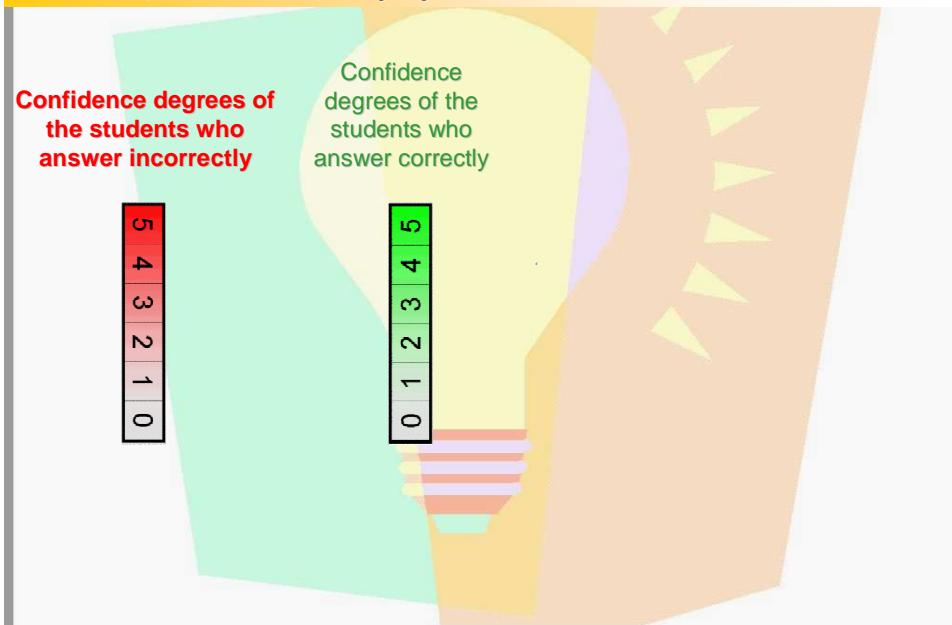
EARLI 11th Biennial conference 2005 - Development of edumetrical indices for the analysis of the spectral quality of higher education standardized tests - Jean-Luc Gilles 5

**Starting intuition for the new indices used for highlighting problematic MCQs and, within those, the proposals which contain anomalies:**



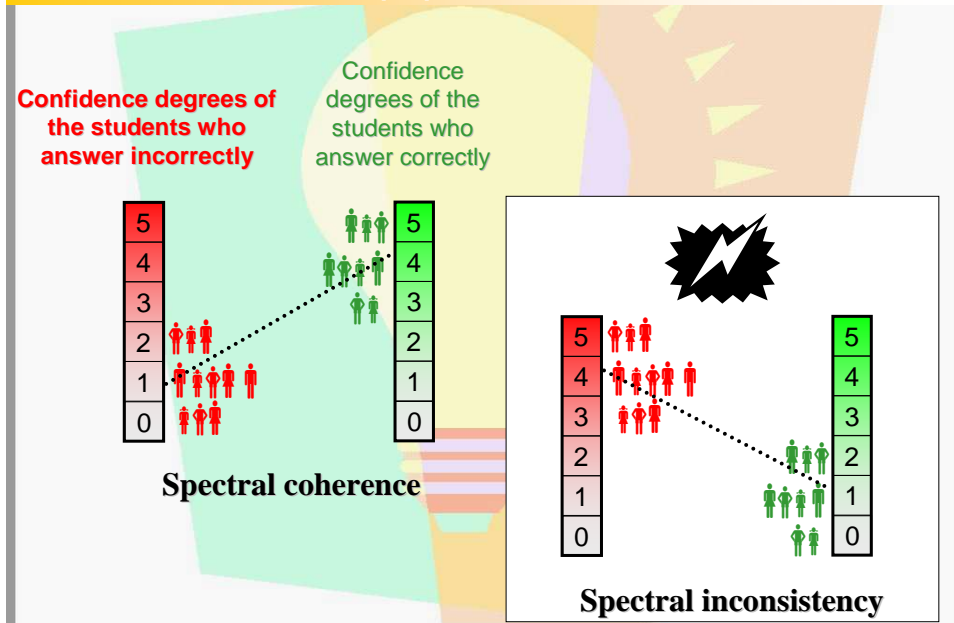
*EARLI 11th Biennial conference 2005 - Development of edumetrical indices for the analysis of the spectral quality of higher education standardized tests - Jean-Luc Gilles 6*

**Starting intuition for the new indices used for highlighting problematic MCQs and, within those, the proposals which contain anomalies:**

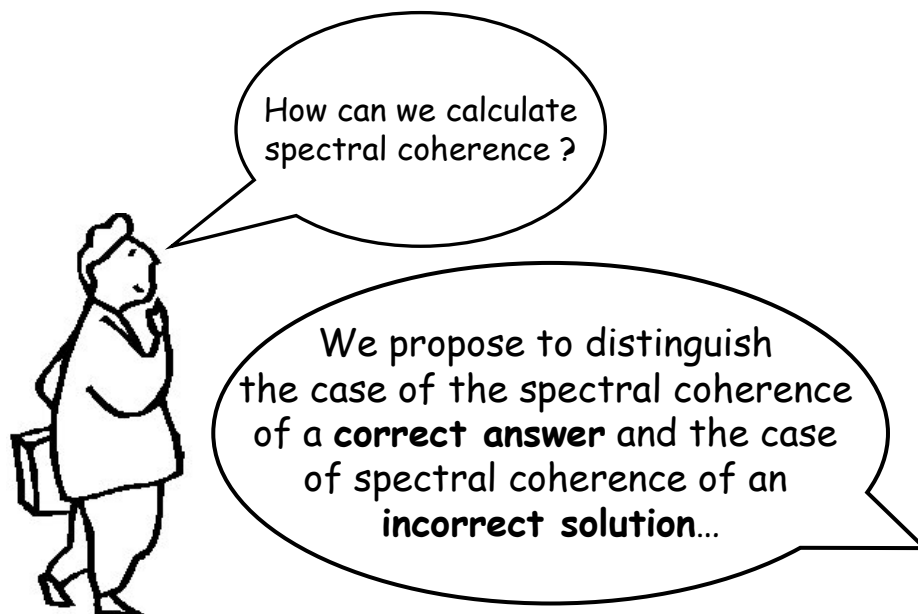


*EARLI 11th Biennial conference 2005 - Development of edumetrical indices for the analysis of the spectral quality of higher education standardized tests - Jean-Luc Gilles 7*

**Starting intuition for the new indices used for highlighting problematic MCQs and, within those, the proposals which contain anomalies:**



EARLI 11th Biennial conference 2005 - Development of edumetrical indices for the analysis of the spectral quality of higher education standardized tests - Jean-Luc Gilles 8



EARLI 11th Biennial conference 2005 - Development of edumetrical indices for the analysis of the spectral quality of higher education standardized tests - Jean-Luc Gilles 9

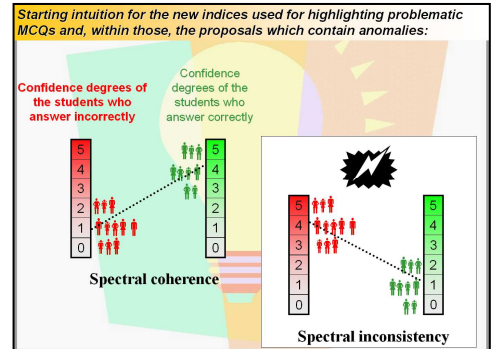
## Spectral contrasted point biserial correlation for the correct solution:

$r_{pb} SC^c$

	OM	S1	S2	S3	S4	CD
Student 1	0	0	0	0	1	0
Student 2	0	1	0	0	0	1
Student 3	0	0	0	1	0	4
Student 4	0	0	0	1	0	5
Student 5	1	0	0	0	0	0
Student 6	0	0	0	1	0	3
Student 7	1	0	0	0	0	0
Student 8	0	0	0	0	1	3
Student 9	0	0	0	1	0	2
Student 10	0	0	1	0	0	0
Student 11	0	0	0	0	1	1
Student 12	0	1	0	0	0	2
Student 13	0	0	0	1	0	5
Student 14	0	1	0	0	0	1
Student 15	0	0	1	0	0	0
Student 16	0	0	1	0	0	3
Student 17	0	0	1	0	0	2
Student 18	0	0	0	0	1	2
Student 19	0	1	0	0	0	0
Student 20	0	0	0	1	0	4

Complete:  
 He ..... a heart attack while he ..... for the bus.  
 1. was having / was waiting  
 2. had / waited  
 3. had / was waiting ← ..... **Correct answer**  
 4. was having / waited

Confidence Degrees (CD) of the students.



analysis of the spectral quality of higher education standardized tests – Jean-Luc Gilles 10

## Spectral contrasted point biserial correlation for the correct solution:

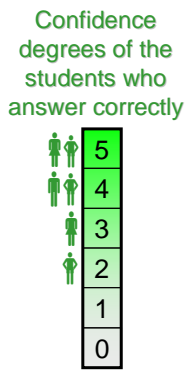
$r_{pb} SC^c$

Which are the confidence degrees of the students who chose the correct solution "S3"?

	S3	CD
Student 1	0	0
Student 2	0	1
Student 3	1	4
Student 4	1	5
Student 5	0	0
Student 6	1	3
Student 7	0	0
Student 8	0	3
Student 9	1	2
Student 10	0	0
Student 11	0	1
Student 12	0	2
Student 13	1	5
Student 14	0	1
Student 15	0	0
Student 16	0	3
Student 17	0	2
Student 18	0	2
Student 19	0	0
Student 20	1	4

Complete:  
 He ..... a heart attack while he ..... for the bus.  
 1. was having / was waiting  
 2. had / waited  
 3. had / was waiting ← ..... **Correct answer**  
 4. was having / waited

Confidence Degrees (CD) of the students.



analysis of the spectral quality of higher education standardized tests – Jean-Luc Gilles 11

## Spectral contrasted point biserial correlation for the correct solution:

$$r_{pb} SC^c$$

Which are the confidence degrees of the students who answer incorrectly?

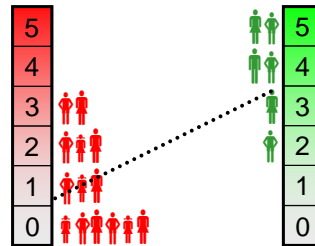
Student	Answer	CD
Student 1	0	0
Student 2	0	1
Student 3	1	4
Student 4	1	5
Student 5	0	0
Student 6	1	3
Student 7	0	0
Student 8	0	3
Student 9	1	2
Student 10	0	0
Student 11	0	1
Student 12	0	2
Student 13	1	5
Student 14	0	1
Student 15	0	0
Student 16	0	3
Student 17	0	2
Student 18	0	2
Student 19	0	0
Student 20	1	4

Complete:  
 He ..... a heart attack while he ..... for the bus.  
 1. was having / was waiting  
 2. had / waited  
 3. had / was waiting ← **Correct answer**  
 4. was having / waited

Confidence Degrees (CD) of the students.

Confidence degrees of the students who answer incorrectly

Confidence degrees of the students who answer correctly



**Spectral contrasted point biserial correlation for the correct solution:**

$$r_{pb} SC^C$$

$$r_{pb} SC^C = \frac{Dx^C - Da^C}{\sigma} \sqrt{pq}$$

with

$Dx^C$  = the average of the confidence Degrees of the students who chose the correct solution

$Da^C$  = the average of the confidence Degrees of the other students who did not choose the correct solution

$\sigma$  = the standard deviation of the confidence degrees of all the students

$p$  = the proportion of students who chose the correct solution of the question

$q$  = the proportion of students who did not choose the correct solution of the question ( $q = 1 - p$ )

EARLI 11th Biennial conference 2005 - Development of edumetrical indices for the analysis of the spectral quality of higher education standardized tests - Jean-Luc Gilles 13

**Spectral contrasted point biserial correlation for the correct solution:**

$$r_{pb} SC^C$$

Complete:  
 He ..... a heart attack while he ..... for the bus.  
 1. was having / was waiting  
 2. had / waited  
 3. had / was waiting ← **Correct answer**  
 4. was having / waited

Confidence Degrees (CD) of the students.

Student	CD
Student 1	0
Student 2	1
Student 3	4
Student 4	5
Student 5	0
Student 6	3
Student 7	0
Student 8	3
Student 9	2
Student 10	0
Student 11	1
Student 12	2
Student 13	5
Student 14	1
Student 15	0
Student 16	3
Student 17	2
Student 18	2
Student 19	0
Student 20	4

By applying the formula:

$$r_{pb} SC^C = \frac{Dx^c - Da^c}{\sigma} \sqrt{pq}$$

$r_{pb} SC^C$  for the correct answer « S3 » = 0,76

The results obtained for the correct solution "S3" are coherent from a spectral point of view

analysis of the spectral quality of higher education standardized tests - Jean-Luc Gilles 14



EARLI 11th Biennial conference 2005 - Development of edumetrical indices for the analysis of the spectral quality of higher education standardized tests - Jean-Luc Gilles 15

**Spectral contrasted point biserial correlation for an incorrect solution:**

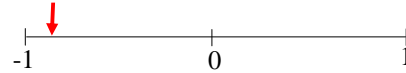
$r_{pb} SC^I$

	Incorrect solution S1	Correct solution S3	CD
Student 2	1	0	1
Student 3	0	1	4
Student 4	0	1	5
Student 6	0	1	3
Student 9	0	1	2
Student 12	1	0	2
Student 13	0	1	5
Student 14	1	0	1
Student 19	1	0	0
Student 20	0	1	4

**Contrasted Treatment**  
 =  
**Elimination of the**  
 « background noise »  
 generated by the data of  
 the other incorrect answers

After contrasted treatment, we calculate the point biserial correlation using the columns « S1 » and « CD »:

$r_{pb} SC^I$  of « S1 »  
 = -0.83



... for the analysis of the spectral quality of higher education standardized tests - Jean-Luc Gilles 16



EARLI 11th Biennial conference 2005 - Development of edumetrical indices for the analysis of the spectral quality of higher education standardized tests - Jean-Luc Gilles 17

## Turbo analysis principle and realism calculation for the Spectral Contrasted point biserial correlation ( $r_{pb\text{SCT}}$ ):

Complete:  
 He ..... a heart attack while he ..... for the bus.

1. was having / was waiting
2. had / waited
3. had / was waiting ← **Correct answer**
4. was having / waited

We calculate the index of Realism ( $R_s$ ) for each student

	S3	CD	REALISM ( $R_s$ )
Student 1	0	0	55
Student 2	0	1	86
Student 3	1	4	81
Student 4	1	5	92
Student 5	0	0	83
Student 6	1	3	74
Student 7	0	0	43
Student 8	0	3	76
Student 9	1	2	43
Student 10	0	0	93
Student 11	0	1	82
Student 12	0	2	66
Student 13	1	5	89
Student 14	0	1	82
Student 15	0	0	94
Student 16	0	3	34
Student 17	0	2	72
Student 18	0	2	87
Student 19	0	0	86
Student 20	1	4	84

← Very good realism!  
← Very bad realism!

of the spectral quality of higher education standardized tests - Jean-Luc Gilles 18

## Turbo analysis principle and realism calculation for the Spectral Contrasted point biserial correlation ( $r_{pb\text{SCT}}$ ):

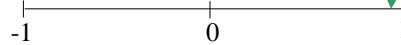
**Turbo analysis = Increasing the reliability of information related to the spectral indices by restricting the data to the students who make less errors in their self-assessments**

Elimination of the data of each student who obtain a realism lower than 80...

	S3	CD	REALISM (Rs)
Student 1	0	0	55
Student 2	0	1	86
Student 3	1	4	81
Student 4	1	5	92
Student 5	0	0	83
Student 6	1	3	74
Student 7	0	0	43
Student 8	0	3	76
Student 9	1	2	43
Student 10	0	0	93
Student 11	0	1	82
Student 12	0	2	66
Student 13	1	5	89
Student 14	0	1	82
Student 15	0	0	94
Student 16	0	3	34
Student 17	0	2	72
Student 18	0	2	87
Student 19	0	0	86
Student 20	1	4	84

...Calculation of the point biserial Spectral Contrasted correlation

$$r_{pb\text{SCT}80} \text{ of } \ll S3 \gg = 0,94$$



of the spectral quality of higher education standardized tests – Jean-Luc Gilles 19



EARLI 11th Biennial conference 2005 - Development of edumetrical indices for the analysis of the spectral quality of higher education standardized tests – Jean-Luc Gilles 20

### Full-size test

We test these new Spectral Contrasted point biserial correlations indices with the data of several thousands of answers and confidence degrees collected during the ten standardized tests of the MOHICAN project supported by the French speaking community of Belgium.

The aim of the MOHICAN project was to gather reliable data on the knowledge of the students entering university.

(Leclercq & al, 2003)

EARLI 11th Biennial conference 2005 - Development of edumetrical indices for the analysis of the spectral quality of higher education standardized tests – Jean-Luc Gilles 21

### Item bank of the “MOHICAN” project :

- Vocabulary (45)
- Syntax and logical structuring of ideas (12)
- Understanding of texts (6)
- Reading of maps, graphics and tables in geography (10)
- Mathematics (22)
- Physics (10)
- Chemistry (8)
- Biology (10)
- General knowledge in arts and culture (25)
- General knowledge in History and Socio-economy (25)

EARLI 11th Biennial conference 2005 - Development of edumetrical indices for the analysis of the spectral quality of higher education standardized tests – Jean-Luc Gilles 22

**Population:**

The academic institutions who participated:

- the Catholic University of Louvain (UCL);
- the Free University of Brussels (ULB),
- the University of Liège (ULg);
- the Academic Faculties Notre-Dame de la Paix in Namur;
- the University of Mons-Hainaut (UMH);
- the Catholic Academic Faculties of Mons (FUCAM);
- the Saint-Louis Academic Faculties in Brussels (FUSL);
- the Polytechnics Faculties of Mons (FPM) and the Agronomics
- the Academic Faculties of Gembloux (FUSAG).

*EARLI 11th Biennial conference 2005 - Development of edumetrical indices for the analysis of the spectral quality of higher education standardized tests – Jean-Luc Gilles 23*

**Population:**

	<i>n</i>
Vocabulary test (VOC)	3,846
Syntax and logical structuring of ideas test (SYN)	3,739
Text comprehension test (COM)	3,420
Reading of maps, graphics and tables in geography test (GEO)	3,688
	<i>n</i>
Mathematics test (MAT)	2,516
Physics test (PHY)	2,497
Chemistry test (CHE)	2,501
Biology test (BIO)	2,507
	<i>n</i>
General knowledge in arts and culture test (ART)	1,392
General knowledge in History and Socio-economy test (HIS)	1,410

*EARLI 11th Biennial conference 2005 - Development of edumetrical indices for the analysis of the spectral quality of higher education standardized tests – Jean-Luc Gilles 24*



EARLI 11th Biennial conference 2005 - Development of edumetrical indices for the analysis of the spectral quality of higher education standardized tests - Jean-Luc Gilles 25

**Full-size test results:  
 The sizes of the populations at 10 turbo analysis levels**

*Size of the 10 tests' population (VOC, SYN, GEO, HIS, ART, MAT, BIO, CHE and PHY) at 10 levels of turbo analysis (from T0 to T90 with a step of 10)*

	VOC	SYN	COM	GEO	HIS	ART	MAT	BIO	CHE	PHY
	<i>n</i>	<i>n</i>	<i>n</i>	<i>n</i>	<i>n</i>	<i>n</i>	<i>n</i>	<i>n</i>	<i>n</i>	<i>n</i>
T0	3.846	3.739	3.420	3.685	1.410	1.392	2.516	2.507	2.501	2.497
T10	3.846	3.734	3.413	3.683	1.409	1.392	2.514	2.507	2.500	2.494
T20	3.843	3.720	3.394	3.678	1.407	1.391	2.499	2.497	2.482	2.472
T30	3.838	3.699	3.348	3.649	1.405	1.387	2.465	2.485	2.461	2.458
T40	3.815	3.658	3.208	3.600	1.399	1.380	2.434	2.442	2.419	2.394
T50	3.773	3.553	2.860	3.474	1.391	1.363	2.398	2.340	2.341	2.342
T60	3.602	3.318	2.279	3.116	1.340	1.323	2.278	2.036	2.096	2.129
T70	3.089	2.691	1.481	2.303	1.101	1.137	1.956	1.511	1.661	1.724
T80	1.806	1.611	756	1.162	584	681	1.276	757	909	1.007
T90	350	439	251	253	73	124	339	151	232	275

EARLI 11th Biennial conference 2005 - Development of edumetrical indices for the analysis of the spectral quality of higher education standardized tests - Jean-Luc Gilles 26

**Full-size test results:**

**Comparison of the  $r_{pb}$ ,  $r_{pb SC}$  and  $r_{pb SCT80}$  values obtained by a particularly discriminating question**

*Values obtained by the 36th question of the vocabulary test*

	OM	P1	P2	P3
$n$	157	* 2.080	323	1.252
%	4	* 54	8	33
$r_{pb}$ →	-0.27	* 0.51	-0.16	-0.30
$r_{pb SC}$ →	xxx	* 0.41	-0.22	-0.35
$n T80$	103	* 1.199	135	612
% T80	5	* 58	7	30
$r_{pb SCT80}$ →	xxx	* 0.58	-0.28	-0.47

Note. OM = omission; P1, P2 and P3 = MCQ's solutions. Values obtained by the correct solution are preceded by a star (\*).

EARLI 11th Biennial conference 2005 - Development of edumetrical indices for the analysis of the spectral quality of higher education standardized tests - Jean-Luc Gilles 27

**Full-size test results:**

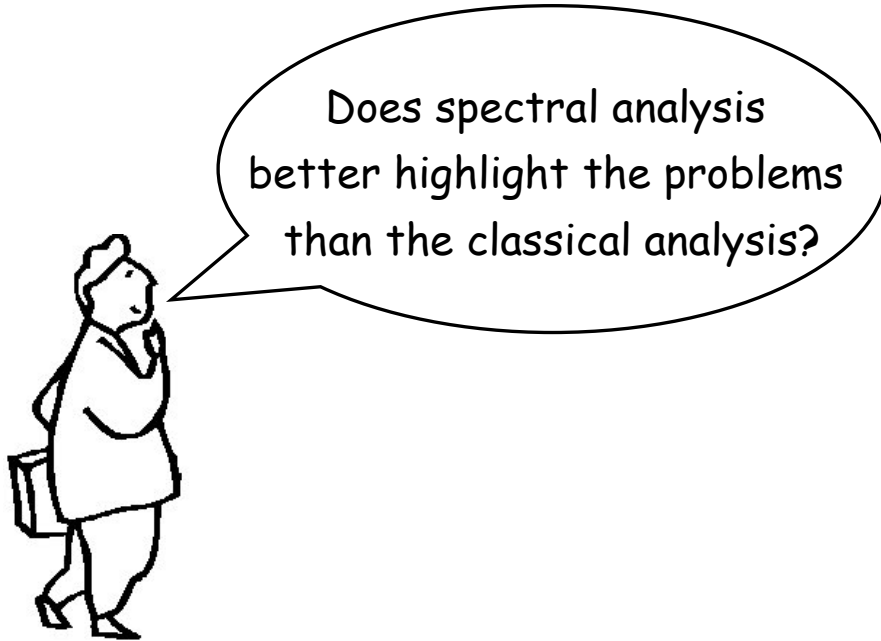
**Comparison of the  $r_{pb}$ ,  $r_{pb SC}$  and  $r_{pb SCT80}$  values obtained by a particularly problematic question**

*Values obtained by the 3<sup>rd</sup> question of the General knowledge in History and Socio-economy test*

	OM	P1	P2	P3	P4	P5	P6
$n$	90	501	132	197	104	258	* 128
%	6	36	9	14	7	18	* 9
$r_{pb}$ →	-0.21	0.36	-0.02	-0.10	-0.06	-0.17	* 0.01
$r_{pb SC}$ →	xxx	0.49	0.05	0.07	0.04	0.19	* -0.15
$n T80$	55	218	67	97	52	111	* 54
% T80	8	33	10	15	8	17	* 8
$r_{pb SCT80}$ →	xxx	0.38	0.00	0.02	-0.03	0.06	* -0.06

Note. OM = omission; P1 to P6 = MCQ's solutions. Values obtained by the correct solution are preceded by a star (\*).

EARLI 11th Biennial conference 2005 - Development of edumetrical indices for the analysis of the spectral quality of higher education standardized tests - Jean-Luc Gilles 28



EARLI 11th Biennial conference 2005 - Development of edumetrical indices for the analysis of the spectral quality of higher education standardized tests – Jean-Luc Gilles 29

## CONCLUSIONS:

### Summary of the full-test results:

- The 10 MOHICAN tests = 173 MCQs
- 16 MCQs presents abnormal values for the classical and/or the spectral correlations
- The experts point 5 MCQs with real problems

#### Concerning the DETECTION of the 5 MCQs highlighted by the experts:

Classical correlation ( $r_{pb}$ )	<i>Detects 2 problematic MCQs out of 5.</i>
New spectral correlation ( $r_{pb\ SC}$ )	<i>Detects all the 5 MCQs highlighted by the experts.</i>
New spectral correlation with turbo analysis ( $r_{pb\ SCT80}$ )	<i>Detects 2 problematic MCQs out of 5.</i>

#### Concerning the FALSE ALARMS (detection of problem whereas the experts do not detect a particular anomaly):

Classical correlation ( $r_{pb}$ )	<i>Activates 10 false alarms.</i>
New spectral correlation ( $r_{pb\ SC}$ )	<i>Activates 7 false alarms.</i>
New spectral correlation with turbo analysis ( $r_{pb\ SCT80}$ )	<i>Activates 1 false alarm.</i>

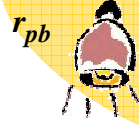
EARLI 11th Biennial conference 2005 - Development of edumetrical indices for the analysis of the spectral quality of higher education standardized tests – Jean-Luc Gilles 30

**CONCLUSIONS:**

*Qualitative analysis: posterior opinions of the experts*



*Classical analysis of discrimination:*



*Spectral analysis:*

$r_{pb}^{SC}$  &  $r_{pb}^{SCT}$



- ① From the point of view of « **detection** »:
  - ▶  $r_{pb}^{SC}$  is more effective
- ② From the point of view of « **less false alarms** »:
  - ▶  $r_{pb}^{SCT80}$  is more effective
  - ▶  $r_{pb}^{SC}$  is more effective than the classical  $r_{pb}$
- ③ Spectral analysis is « **test free** »

EARLI 11th Biennial conference 2005 - Development of edumetrical indices for the analysis of the spectral quality of higher education standardized tests – Jean-Luc Gilles 31

Thank you for your attention!

For more information concerning the spectral quality analysis of standardized tests,

do not hesitate to contact me:

[jl.gilles@ulg.ac.be](mailto:jl.gilles@ulg.ac.be)

<http://www.smart.ulg.ac.be/earli-2005/spectral-analysis.ppt>

EARLI 11th Biennial conference 2005 - Development of edumetrical indices for the analysis of the spectral quality of higher education standardized tests – Jean-Luc Gilles 32

**Annexe III : Convention de recherche First Spin Off « SAFES »**

Référence ARD : R.12.07
Suivi par : R. Abouk
Adm. Recherche & Développement

Conv. First Spin-off – n°516157 - SAFES - 2005-09-13

- 1

**Convention n° 516157**

**Subvention FIRST - Spin off - SAFES**

***Développement d'une méthode d'audit, d'une typologie de services et d'un module de labellisation***

**Entre**

La Région wallonne,  
Représentée par Madame Marie-Dominique Simonet,  
Ministre de la Recherche et des Technologies nouvelles et des Relations extérieures,  
Rue des Brigades d'Irlande, 2 à 5000 NAMUR,  
Ci-après dénommée la REGION,

**Et**

L'Université de Liège  
Place du 20 Août, 7 à 4000 Liège  
Représenté(e) par Monsieur B. Rentier, Recteur,  
Ci-après dénommé(e) l'UNITE,

**Il est exposé ce qui suit :**

Par l'arrêté du \_\_\_\_\_, la REGION a accordé à l'UNITE une subvention d'un montant maximal de 111380 euro, à charge de l'article 61.01.02 de la division organique 12, programme 12.02, titre II du budget des dépenses de la Région wallonne pour l'année budgétaire 2005.

Cet arrêté dispose, en substance, que les relations qui en découlent font l'objet d'une convention.

## Par conséquent, il est convenu ce qui suit :

### 1. Définitions

Au sens de la présente convention, on entend par :

- « ARRETE D'OCTROI » : l'arrêté, mentionné ci-avant, par lequel la REGION a accordé une subvention à l'UNITE;
- « ARRETE REGLEMENTAIRE » : l'arrêté du Gouvernement wallon du 29 septembre 1994 relatif aux aides et interventions pour la recherche et les technologies;
- « DECRET » : le décret du 5 juillet 1990 relatif aux aides et aux interventions de la Région wallonne pour la recherche et les technologies, modifié par le décret-programme du 17 décembre 1997 portant diverses mesures en matière d'impôts, taxes et redevances, de logement, de recherche, d'environnement, de pouvoirs locaux et de transports;
- « PROGRAMME » : le programme dit *FIRST (Formation et impulsion à la recherche scientifique et technologique) - Spin off*, qui vise à inciter un chercheur universitaire à étudier les conditions d'exploitation industrielle et commerciale des résultats de ses recherches et à être, si possible, à la base de la création d'une entreprise sur le territoire de la REGION ;
- « SUBVENTION » : la subvention visée par l'ARRETE D'OCTROI;
- « DEPENSES ADMISSIBLES » : les dépenses que la SUBVENTION est appelée à couvrir, sans préjudice des vérifications visées à l'article 7;
- « TRAVAUX » : les activités de recherche industrielle de base qui font l'objet de la SUBVENTION;
- « CHERCHEUR » : le chercheur que l'UNITE engage au titre du PROGRAMME;
- « RESULTATS » : l'ensemble des résultats obtenus au fur et à mesure de la réalisation des TRAVAUX et des résultats obtenus à la fin de celle-ci, constitué notamment des rapports, des plans, des notes de calcul, des cahiers de laboratoire, du savoir-faire, des installations pilotes, des codes sources, de l'ensemble des fichiers nécessaires à la création de code exécutable;
- « DIRECTION GENERALE » : la Direction générale des Technologies, de la Recherche et de l'Energie du Ministère de la Région wallonne.
- « FORMATION » : la formation en gestion et en création d'entreprises que le chercheur s'engage à suivre dans un établissement reconnu par la Communauté française de Belgique ;
- « PARRAIN » : expert proposé par l'UNITE qui est à la fois compétent sur les plans de la technique et du management.

### 2. Objet de la présente convention

- 2.1. La présente convention règle les relations réciproques de la REGION et de l'UNITE concernant l'engagement du CHERCHEUR, ses prestations et la réalisation des TRAVAUX, sans préjudice :
  - du DECRET;
  - de l'ARRETE REGLEMENTAIRE;
  - des dispositions légales et réglementaires relatives à la comptabilité et aux aides publiques, notamment les articles 55 à 58 des lois sur la comptabilité de l'Etat, coordonnées le 17 juillet 1991;

- des dispositions légales et réglementaires relatives au fonctionnement du Gouvernement wallon et de ses services, notamment l'arrêté du Gouvernement wallon du 14 octobre 1993 relatif aux délégations de pouvoirs spécifiques au Ministère de la Région wallonne.
- 2.2 Les TRAVAUX auront pour objet la mise au point d'un produit, procédé, ou service nouveau. Il comprendront également l'élaboration d'un schéma crédible de valorisation contenant une étude de faisabilité technico-économique de l'exploitation des RESULTATS.
- 2.3. La présente convention relève du PROGRAMME. Les deux parties s'engagent à respecter les directives d'application de celui-ci, qui constituent l'annexe n° 1 de la présente convention.
- 2.4. Dans les six mois qui suivent la date de signature de la présente convention, l'UNITE engage le CHERCHEUR dans les liens d'un contrat de travail à durée déterminée, l'affecte à la réalisation des TRAVAUX et lui fait signer un accord de confidentialité portant sur ceux-ci.  
  
Dès cet engagement, l'UNITE adresse à la REGION :
  - une copie du contrat de travail et de l'accord de confidentialité;
  - le curriculum vitae du CHERCHEUR;
  - une déclaration par laquelle celui-ci reconnaît avoir pris connaissance de la présente convention et déclare adhérer à ses dispositions;
  - une attestation d'inscription à la FORMATION, ou, le cas échéant, de réussite d'une formation en gestion et en création d'entreprises.
- 2.5. Les TRAVAUX sont réalisés sous la direction scientifique et technique de l'UNITE.
- 2.6. Les articles 14 et suivants constituent les stipulations particulières de la présente convention. Ils comportent notamment :
  - une description des TRAVAUX, de leurs objectifs scientifiques, techniques, industriels et commerciaux et des lieux où ils seront réalisés;
  - le calendrier prévisionnel des TRAVAUX et des prestations du CHERCHEUR;
  - le budget prévisionnel des DEPENSES ADMISSIBLES.

### **3. Organisation des relations entre les parties**

- 3.1. L'UNITE adresse la correspondance relative à l'exécution de la présente convention et destinée à la REGION au Directeur général de la DIRECTION GENERALE, avenue Prince de Liège, 7 à 5100 Jambes.

La REGION adresse la correspondance relative à l'exécution de la présente convention et destinée à l'UNITE au Chef de projet désigné à l'article 18.

- 3.2. Tout délai exprimé en mois dans la présente convention se calcule de quantième à veille de quantième, depuis le lendemain du jour de l'acte ou de l'événement qui y donne cours.

Au sens de la présente convention, les jours ouvrables sont les jours autres que les samedis, dimanches et jours fériés légaux.

#### **4. Modalités générales des prestations du CHERCHEUR**

- 4.1. L'UNITE a seule qualité d'employeur à l'égard du CHERCHEUR. Elle veille en tout temps à respecter les termes de son contrat de travail et à lui permettre d'acquérir une expérience professionnelle conforme aux objectifs du PROGRAMME.
- 4.2. L'UNITE s'engage à procurer une aide scientifique et technique à la réalisation des TRAVAUX, de la manière définie à l'article 15.
- 4.3. Le CHERCHEUR n'est pas tenu de réaliser des prestations qui n'auraient manifestement aucun rapport avec l'économie générale du PROGRAMME. Il informe la REGION chaque fois que l'UNITE méconnaît cette disposition.
- 4.4. Pendant toute la durée des TRAVAUX, le CHERCHEUR s'interdit d'exercer toute activité professionnelle autre que celle qui est prévue dans la présente convention.
- 4.5. Le CHERCHEUR suivra une FORMATION. Sauf accord des parties à la présente, elle ne pourra dépasser 30 heures par mois, à moins d'être suivie en horaire décalé. Les épreuves de la FORMATION devront être présentées et réussies pendant la durée des TRAVAUX. Sera dispensé de la FORMATION, le CHERCHEUR qui est détenteur d'un titre considéré comme équivalent par les institutions universitaires et les hautes écoles de la Communauté française de Belgique.
- 4.6. Le CHERCHEUR sera aidé par un PARRAIN proposé au choix du Comité d'accompagnement du PROGRAMME.

#### **5. Rapports**

- 5.1. Les rapports d'activités et les rapports scientifiques et techniques visés à l'article 7 de l'ARRETE REGLEMENTAIRE ont une fréquence semestrielle, correspondant aux périodes de six mois qui s'écoulent à partir de la date de signature du contrat de travail du CHERCHEUR. L'UNITE rédige et adresse les rapports à la REGION dans les trente jours ouvrables qui suivent la période à laquelle ils se rapportent.

Chaque rapport d'activités comprend :

- un exposé de l'état d'avancement des TRAVAUX, accompagné d'un calendrier d'exécution;
- un exposé des activités du chercheur ayant pour objet la création d'une société nouvelle.

Conv. First Spin-off – n °516157 - SAFES - 2005-09-13

- 5

- une évaluation des prestations du CHERCHEUR, contresignée et annotée par l'intéressé, qui expose notamment l'expérience professionnelle qu'il estime avoir acquise au cours du semestre.
- Celui-ci est rédigé selon le modèle disponible sur le site web de la REGION à l'adresse : <http://www.Recherche-technologie.wallonie.be>

Chaque rapport scientifique et technique comprend :

- un exposé des RESULTATS atteints;
- une évaluation de leurs possibilités de valorisation industrielle ou commerciale.

- 5.2. Dans les trente jours ouvrables qui suivent la fin des TRAVAUX, l'UNITE rédige et adresse à la REGION un rapport d'activités et un rapport scientifique et technique finals.

Le rapport d'activités final comprend :

- un exposé récapitulatif du déroulement des TRAVAUX, accompagné du calendrier d'exécution;
  - une évaluation des prestations du CHERCHEUR, contresignée et annotée par l'intéressé, qui expose notamment l'expérience professionnelle qu'il estime avoir acquise tout au long des TRAVAUX;
  - l'attestation de réussite de la FORMATION, le cas échéant.
- Il est rédigé selon le modèle cité au point 5.1.

Le rapport scientifique et technique final comprend :

- un exposé détaillé des RESULTATS atteints à la fin des TRAVAUX;
- une évaluation de leurs possibilités de valorisation industrielle ou commerciale comprenant un plan d'affaires sur 5 ans et une évaluation des potentialités du marché visé.

- 5.3. Un mois après la fin des TRAVAUX, l'UNITE adresse à la REGION, un rapport d'utilisation visé à l'article 7 de l'ARRETE REGLEMENTAIRE, qui comprend :
- les éléments prouvant le respect des articles 8.3 et 8.4;
  - un exposé des retombées industrielles ou commerciales, effectives ou potentielles, des RESULTATS en Wallonie, ainsi que la grille d'évaluation dont le modèle peut être téléchargé sur le site web de la REGION : <http://www.Recherche-technologie.wallonie.be>;
  - un résumé des prestations du CHERCHEUR depuis la fin des TRAVAUX.
- 5.4. L'UNITE répond, de la manière la plus complète et la plus rapide possible, à toute demande d'information de la REGION relative à l'exécution de la présente convention, dans la mesure où la demande n'entraîne pas une charge de travail anormale ou des frais excessifs.

## **6. Budget des DEPENSES ADMISSIBLES**

- 6.1. Les DEPENSES ADMISSIBLES sont limitées à celles que visent les articles 6.2. et 6.3., pour autant qu'elles soient relatives à la période de réalisation des TRAVAUX au sens de l'article 13. Elles sont appelées à être financées en totalité par la

## SUBVENTION.

Le budget prévisionnel des DEPENSES ADMISSIBLES figure à l'article 16.

Il peut notamment contribuer au financement d'une thèse de doctorat ou de post doctorat.

- 6.2. Les DEPENSES ADMISSIBLES de personnel ne portent que sur les éléments de salaire du CHERCHEUR, qui comprennent :
- la rémunération brute indexée, plafonnée suivant le barème en vigueur dans les universités publiques;
  - les charges sociales patronales;
  - les assurances légales;
  - les indemnités et allocations dues en vertu de dispositions légales et réglementaires et de conventions collectives de travail;
  - les frais d'inscription à la FORMATION.

Sous réserve du respect du barème visé à l'alinéa premier, le CHERCHEUR bénéficie de conditions salariales identiques à celles que pratique habituellement l'UNITE pour le personnel de même niveau de responsabilité, de qualification et d'ancienneté.

- 6.2. Les DEPENSES ADMISSIBLES de fonctionnement sont établies de manière forfaitaire à 5.000 € par semestre.

Le montant des dépenses admissibles du personnel est établi au moment de la signature de la CONVENTION, sur base d'éléments budgétaires fournis par l'UNITE. Aucun avenant d'octroi d'un complément budgétaire ne sera accordé par la Région après la date de signature de la CONVENTION.

## 7. Relevés des dépenses et paiements

- 7.1. Dans les quinze jours ouvrables de la notification à l'UNITE de l'ARRETE D'OCTROI, la REGION met en liquidation un fonds de roulement dont le montant est déterminé à l'article 18.
- 7.2. A la réception d'un rapport d'activités et d'un rapport scientifique et technique visés à l'article 5.1., la REGION vérifie si les éléments exposés sont complets et conformes aux stipulations de la présente convention; si ce n'est pas le cas, elle demande des informations complémentaires à l'UNITE dans les trente jours ouvrables de la réception des rapports, faute de quoi les éléments sont réputés complets et conformes à l'expiration de ce délai.

La REGION vérifie également le relevé des dépenses, la déclaration de créance et les pièces justificatives correspondantes qui accompagnent le rapport d'activités.

La présomption de l'alinéa premier ne porte en aucune façon sur les dépenses admises, dont la REGION détermine le montant. Elle met celui-ci en liquidation, de façon à

Conv. First Spin-off – n °516157 - SAFES - 2005-09-13

- 7

reconstituer le fonds de roulement visé à l'article 7.1., pour autant que les rapports soient complets et conformes, ou réputés tels.

- 7.3. La REGION procède de la manière visée à l'article 7.2. jusqu'à ce que le solde de la SUBVENTION restant à liquider soit réduit à 10 % (dix pour cent) de celle-ci.
- 7.4. A la réception du rapport d'activités et du rapport scientifique et technique finals visés à l'article 5.2., la REGION vérifie si les éléments exposés sont complets et conformes aux stipulations de la présente convention; si ce n'est pas le cas, elle demande des informations complémentaires à l'UNITE dans les trente jours ouvrables de la réception des rapports, faute de quoi les éléments sont réputés complets et conformes à l'expiration de ce délai.

La REGION vérifie également le relevé des dépenses, la déclaration de créance et les pièces justificatives correspondantes qui accompagnent le rapport d'activités final.

La présomption de l'alinéa premier ne porte en aucune façon sur les dépenses admises, dont la REGION détermine le montant. Elle met le solde de la SUBVENTION en liquidation, pour autant que les rapports finals soient complets et conformes, ou réputés tels.

- 7.5. La part des DEPENSES ADMISSIBLES qui est ou qui serait financée par un organisme public belge, étranger ou international ne peut être incluse dans aucune déclaration de créance.
- 7.6. Toute mise en liquidation s'effectue par virement au compte financier de l'UNITE qui figure à l'article 18.

L'alinéa premier ne s'applique pas à une mise en liquidation, visée à l'article 7.2. ou à l'article 7.4., pour laquelle la REGION se voit notifier en temps utile d'autres modalités, par l'UNITE ou de son accord.

## **8. Valorisation des RESULTATS**

- 8.1. Conformément à l'article 3, paragraphe 3, alinéa 3, du DECRET, l'UNITE est propriétaire des RESULTATS.
- 8.2. L'UNITE valorise les RESULTATS en les présentant sous une forme qui les rende accessibles à des tiers et en les proposant en priorité à des entreprises qui les exploiteraient à terme sur le territoire de la Région wallonne.
- 8.3. Avant toute concession ou cession relative aux RESULTATS, l'UNITE accorde un droit de premier refus au CHERCHEUR, pendant la recherche et jusqu'à trois mois après la fin de la recherche.

- 8.4. Dans l'hypothèse où le CHERCHEUR marque la volonté de créer une société commerciale pour exploiter les RESULTATS, l'UNITE lui accorde sur les RESULTATS une licence:
- gratuite pendant 5 ans ; après cette période, des redevances pourront être demandées par l'UNITE ;
  - non cessible à un tiers sans accord préalable de l'UNITE;
  - exclusive, à condition qu'il y ait une exploitation effective. En cas de non exploitation, au delà de la période de gratuité de 5 ans, la licence deviendra non exclusive et pourra même être retirée.
- 8.5. L'UNITE doit informer le CHERCHEUR du statut précis des RESULTATS.
- 8.6. L'UNITE doit informer la REGION de la création d'une société par le CHERCHEUR. Le siège d'exploitation de cette société doit être situé sur le territoire de la REGION , pendant cinq ans au moins.

## 9. Obligation de secret, publications, communications

- 9.1. Les RESULTATS et tous autres documents, informations, connaissances et savoir-faire relatifs aux TRAVAUX sont secrets. Les parties s'engagent à leur conserver ce caractère, dans la mesure compatible avec la valorisation des RESULTATS, telle que visée à l'article 8.

Cette obligation entraîne notamment que chaque partie :

- ne peut faire des éléments couverts par secret que l'usage autorisé par la présente convention;
- ne peut les diffuser auprès de son personnel que dans la mesure strictement nécessaire;
- exige la même obligation de secret de son personnel, de ses sous-traitants et de ses cocontractants.

- 9.2. L'obligation de secret et l'interdiction d'usage stipulées à l'article 9.1. ne s'appliquent pas aux informations dont la partie concernée apporte la preuve :
- qu'elles ont déjà fait l'objet d'une publication à la date de signature de la présente convention, ou
  - qu'elles étaient déjà en sa possession à la même date, ou
  - qu'elles lui ont été communiquées par un tiers sans qu'il y ait à l'origine violation d'une quelconque obligation de secret, ou
  - qu'elles sont tombées dans le domaine public, sauf si cela est dû à une faute quelconque de la partie, d'un membre de son personnel, d'un sous-traitant ou d'un contractant.

Ces exceptions sont de stricte interprétation; elles ne s'étendent qu'aux données explicites et non à leurs développements scientifiques ou techniques, mêmes implicites, obtenus au cours des TRAVAUX.

- 9.3. L'UNITE adresse à la REGION, dès sa parution, toute publication ou communication relative aux TRAVAUX ou aux RESULTATS. Elle veille à ce que la publication ou la communication évoque le financement de la REGION au titre du PROGRAMME.

## 10. Accès de la REGION aux installations de l'UNITE

- 10.1. L'UNITE accepte et facilite l'exercice par la REGION des contrôles administratifs, techniques et scientifiques destinés à vérifier la bonne exécution de la présente convention.

L'UNITE accepte et facilite également les contrôles prévus par les dispositions légales et réglementaires relatives à la comptabilité et aux aides publiques.

- 10.2. A cet effet, les personnes déléguées par la REGION et habilitées, en vertu de l'arrêté royal du 26 avril 1968 réglant l'organisation et la coordination des contrôles de l'octroi et de l'emploi des subventions, pour exercer le contrôle visé par les articles 55 à 58 des lois sur la comptabilité de l'Etat, coordonnées le 17 juillet 1991, ont accès aux lieux où

se déroulent les activités relevant des TRAVAUX et aux lieux où le CHERCHEUR réalise ses prestations.

Ces personnes peuvent prendre connaissance, sur place dans tous les établissements de l'UNITE, de tout document comptable ou autre, utile pour vérifier la bonne exécution de la présente convention.

L'UNITE peut désigner un représentant pour accompagner les personnes déléguées par la REGION.

- 10.3. Les personnes déléguées par la REGION peuvent se faire accompagner d'experts.

La REGION fait contracter aux experts un engagement suivant lequel ils s'interdisent de divulguer à des tiers les informations recueillies à l'occasion de ces contrôles et d'en faire usage.

Ces obligations ne couvrent pas les informations qui sont déjà dans le domaine public au moment du contrôle, ni celles qui y tombent après le contrôle sans faute de l'expert, ni celles dont l'expert apporte la preuve qu'il les détenait légitimement au moment du contrôle, ou qu'il les a reçues postérieurement d'un tiers sans qu'il y ait à l'origine violation d'une quelconque obligation de secret.

## 11. Responsabilité

- 11.1. La REGION ne peut en aucune façon être tenue pour responsable des dommages aux personnes et aux biens résultant directement ou indirectement de l'exécution de la présente convention par l'UNITE.

11.2. La REGION ne peut en aucune façon être tenue pour responsable du traitement comptable et fiscal que l'UNITE réserve à la SUBVENTION.

## **12. Droit applicable et juridictions compétentes**

La présente convention est régie par le droit belge.

Tout litige relatif à son interprétation ou à son exécution est de la compétence des juridictions de Namur.

## **13. Durée**

13.1. Les TRAVAUX sont réputés prendre cours à la date de signature du contrat de travail du CHERCHEUR. Ils sont réputés se terminer vingt-quatre mois après cette date.

13.2. Au plus tard trois mois avant la fin des TRAVAUX au sens de l'article 13.1., l'UNITE peut solliciter leur prolongation pour une période de douze mois en adressant à la REGION une demande qui comporte :

- un exposé des raisons qui motivent la demande de prolongation;
- une description des activités qui constitueraient la poursuite des TRAVAUX;
- le calendrier prévisionnel de la poursuite des TRAVAUX et des prestations du CHERCHEUR;
- le budget prévisionnel correspondant, limité aux DEPENSES ADMISSIBLES que visent les articles 6.2. et 6.3.

Cette demande est rédigée selon le modèle disponible sur le site web de la REGION à l'adresse : <http://www.Recherche-technologie.wallonie.be>

13.3. En cas d'accord sur la prolongation, les parties concluent un avenant à la présente convention.

13.4. A l'issue de cette prolongation, une deuxième prolongation d'un an pourra être demandée par l'UNITE suivant les conditions du point 13.2.

13.5. Il ne peut y avoir aucune interruption entre la fin de la période de deux ans couverte par la présente et les deux éventuelles prolongations d'un an.

13.6. Le poste de CHERCHEUR est nominatif. Tout remplacement du CHERCHEUR en cours de mandat (prolongation éventuelle comprise) est interdit. Si le CHERCHEUR cesse définitivement les prestations prévues dans la présente CONVENTION, celle-ci prend fin anticipativement et la REGION procède à sa clôture.

13.7. En cas de force majeure liée à l'indisponibilité du CHERCHEUR (maladie, maternité, ...), une suspension de maximum 6 mois de la CONVENTION peut être accordée par la REGION. Cette suspension ne sera effective qu'après l'accord octroyé par la REGION.

**Stipulations particulières**

**14. Description des TRAVAUX**

Les préoccupations relatives à la qualité des évaluations standardisées des acquis des formés s'insèrent dans un mouvement qui depuis le début des années '90, touche les institutions belges et européennes (Kaufman & al., 1995 ; Segers & Dochy, 1996). Ce mouvement participe lui-même à une tendance mondiale d'introduction du management de la qualité dans la sphère des activités d'enseignement et de formation (Ramsden, 1991 ; Nightingale & O'Neil, 1994 ; Zink & Schmidt, 1995). Les évaluations standardisées de qualité s'insèrent dans ce contexte en facilitant les processus de cartographie des compétences au sein des entreprises, administrations et institutions, ce qui leur permet de réagir plus efficacement aux demandes du marché. En outre, l'évaluation constitue aussi un enjeu pour les personnes évaluées, notamment lorsque des avantages financiers sont en jeu (tests certificatifs), elle est alors l'objet de toute la vigilance d'organes de régulation et de contrôle (les partenaires sociaux).

Dès lors, les entreprises, administrations et institutions de formation sont de plus en plus sensibles à la nécessité d'apporter la preuve de la qualité (notamment en terme d'équité et donc bien souvent en ayant recours aux tests standardisés) de leurs épreuves certificatives.

Le projet « Smart Advices For Evaluation Systems (SAFES) » a pour objectif de créer une spin-off active dans le secteur des évaluations standardisées de qualité. Cette société proposera des produits / services originaux à forte valeur ajoutée basés sur le développement d'une méthode d'audit, d'une typologie de services et d'un module de labellisation en testing qui viendra se greffer sur la plateforme électronique de construction et de gestion qualité de tests standardisés (projet intitulé electronic Construction and Quality Control in Standardized Testing (e-C&QCST) - programme WIST de la DGTRE).

Dans ce contexte, les méthodes d'audit développées en amont de l'utilisation d'e-C&QCST permettront de pointer les éléments perfectibles dans les systèmes de testing pour ensuite proposer des améliorations via une typologie d'offres de services. Les outils de labellisation permettront quant à eux d'apporter la preuve de la qualité des systèmes de certification des compétences et se situeront en aval d'e-C&QCST. Notre calendrier de travail prévoit la mise à disposition de ces services en même temps que la diffusion de la plateforme début 2007.

**15 Calendrier prévisionnel des TRAVAUX et des prestations du CHERCHEUR**

Voir annexe 2.

**16. Budget prévisionnel des DEPENSES ADMISSIBLES**

Personnel (Dont maximum 1.250 euro pour la formation)	91380 euro
Frais de fonctionnement	20.000 euros
<b>TOTAL</b>	<b>111380 euro</b>



*Annexe 1*

**DIRECTION GENERALE DES TECHNOLOGIES, DE LA RECHERCHE ET DE L'ENERGIE  
PROGRAMME FIRST SPIN-OFF**

**APPEL A PROPOSITIONS  
Exercice 2005  
PRESENTATION GENERALE**

**1. Description générale**

Les objectifs du programme FIRST SPIN-OFF sont :

- Accroître le potentiel scientifique et technologique des unités de recherche universitaire ou de niveau universitaire via la réalisation d'une **recherche** susceptible d'avoir un impact positif sur le développement économique et social de la Région wallonne et sur la création d'emplois durables;
- Former de futurs cadres d'entreprise aux technologies émergentes ainsi qu'aux techniques de création et de management d'activités industrielles;
- Inciter le chercheur universitaire à étudier les conditions d'exploitation industrielle et commerciale des résultats de ses recherches et à être, si possible, à la base d'une création d'entreprise en Région wallonne ;
- Au terme de cet appel, l'objectif est d'octroyer 20 mandats.

**2. Qui peut proposer ?**

- Les universités de la Communauté française disposant d'une implantation en Région wallonne.
- Les centres de recherche, asbl associées aux Hautes Ecoles de la Communauté française dispensant un enseignement supérieur technique de type long (au sens de l'article 2 de la loi du 7 juillet 1970 et de l'article 17 de la loi du 5 août 1995) et disposant d'une implantation en Région wallonne.

Dans le cas où le projet de recherche proposé impliquerait la collaboration de plusieurs unités de recherche universitaires ou de niveau universitaire, un seul promoteur est considéré comme porte-parole du projet auprès de la D.G.T.R.E..

Conv. First Spin-off – n °516157 - SAFES - 2005-09-13

## Annexe 2

### 9. Plan de travail de l'ensemble du projet

Le plan de travail doit tenir compte du rapport d'évaluation sur les potentialités de créer une activité nouvelle via la création d'une spin-off. Celui-ci doit être étoffé tout au long de la recherche en tenant compte notamment des éléments repris au point 3 de l'appel à propositions.

#### 9.1. Description des tâches (1200 mots maximum)

##### Etape 1 : phase exploratoire

##### **Volet Recherche & Développement**

###### *Lignes d'actions « Audit systèmes de testing »*

Comme décrit au point 5.2 du présent projet, l'objectif ici est de développer une méthode d'audit des systèmes de testing originale.

Les étapes du développement :

###### *a. Concevoir les normes à partir de nos modèles scientifiques :*

En effet, les modèles que nous mettons en œuvre sont le fruit de recherches scientifiques qui ont été validées. Ces modèles font apparaître une série d'étapes par lesquelles il convient de se conformer afin de s'assurer un maximum de chances de mettre en place un système de testing de qualité permettant de répondre aux critères de qualité définis plus haut dans le point « 5.1 Contexte scientifique ». De ces étapes, il sera donc possible d'établir des normes auxquelles le système de testing devra répondre.

###### *b. Déterminer le champ d'application de chacune des normes*

Il sera essentiel de définir la portée de ces normes et de définir un champ d'application pour chacune.

###### *c. Définir les indicateurs qui serviront à quantifier le degré d'adéquation des normes avec leur champ d'application respectif.*

###### *d. Etablir un mode de recueil de l'information*

Toute méthode d'audit commence par un recueil d'informations en vue d'établir un état des lieux et donc un premier rapport d'audit.

###### *e. Création d'un avant projet cahier des charges reprenant tous les points de non-conformité aux normes établies.*

Ce cahier des charges fera état des écarts par rapport aux normes établies et proposera les solutions d'amélioration sous forme d'actions correctives ou de propositions de procédures à adopter.

*f. Créer un avant projet des manuels de procédures d'audit qui seront utilisées*

...afin d'assurer un même degré de service, une cohérence claire et pouvoir assurer une transparence pour l'institution demandeuse.

**Ligne d'actions « Processus de labellisation »**

L'objectif est de concevoir l'outil de support qui permettra une prise en charge informatisée de la labellisation du système de testing. Ce support sera un module informatique programmé en PHP MySQL et qui viendra se greffer à la plateforme e-C&QCST.

- a. Conception d'un cahier des charges répondant aux exigences
- b. Programmation du module sur base du cahier des charges
- c. Essai applicatif auprès de partenaires de terrain
- d. Correction et régulation du module

**Volet Transfert de technologie/expertise**

**Ligne d'actions « Typologie d'offres / scénarios d'intervention »**

- a. Répertoire, en amont et en aval des processus de mise en place de système de testing, les besoins (produits/services) en matière d'appropriation des applications d'évaluation
- b. Définir le champs d'application de chacun des produits/ services
- c. Essais applicatifs auprès des partenaires de terrain
- d. Création des supports nécessaires

**Ligne d'actions « Exploration du marché »**

Réalisation d'une première étude de faisabilité de création d'une spin-off qui valorisera des produits et services en matière de testing standardisé.

- a. Réalisation d'une étude de marché comprenant :
  - i. L'analyse de ressources documentaires sur le secteur belge mais aussi étranger actif dans le domaine de l'évaluation standardisée.
  - ii. La collecte d'informations sur les concurrents éventuels (services/produits offerts, prix, stratégie commerciale, positionnement...)

- iii. L'analyse du marché de l'évaluation :
  1. *Evaluation du marché (taille, dimensions du marché, comportements...)*
  2. *Etude qualitative des besoins des acteurs du monde de l'entreprise, de la formation et de l'enseignement*
  3. *Segmentation du marché*
- b. *Etude de marché sur les normes, les agrégations, les certifications et les labels existant en matière de formation :*
  - i. Analyse documentaire du secteur en vue de dégager quelle est la place de l'évaluation dans la réglementation
  - ii. Etude qualitative sur les procédés des acteurs du secteur en vue éventuellement de créer des partenariats ou des liens de coopération
- c. *Elaboration d'un avant projet de stratégie marketing :*
  - i. Ciblage
  - ii. Positionnement des produits et services offerts sur les différents marchés
  - iii. Première ébauche du plan d'affaire
  - iv. Stratégie de communication
    1. *Avant projet des médias de communication*
    2. *Premiers contacts avec des éventuels canaux de diffusion*
  - v. Stratégie de « pricing »
  - vi. Etude financière des coûts engendrés et du return possible suivants le volume d'affaire imaginé à partir d'un premier plan d'affaire

## Etape 2 : Phase opérationnelle

### **Volet Recherche & Développement**

#### *Lignes d'actions « Audit systèmes de testing »*

Le but est ici de valider la pertinence et les procédures mises en œuvre dans le cadre des audits de système de testing.

- a. *Expérimentation des procédures de l'avant projet de manuel d'audit auprès des partenaires de terrain voire de futurs clients potentiels.*
- b. *Evaluation de chacune des procédures en terme qualitatif et quantitatif ainsi que des problèmes rencontrés.*
- c. *Récolte des avis et remarques en vue de la régulation et l'amélioration de la méthode.*
- d. *Création de la version finale du manuel de procédures.*

## *Ligne d'actions « Processus de labellisation »*

L'objectif de cette ligne d'actions est de valider le processus de labellisation d'un système de testing.

- a. *Présentation de la méthode de labellisation aux partenaires de terrain en vue de récolter les avis et remarques*
- b. *Réguler et améliorer la méthode*
- c. *Présentation du module informatisé de labellisation aux partenaires de terrain et essai applicatif réel*
- d. *Récolte d'information sur les avis et remarques ainsi les problèmes techniques rencontrés*
- e. *Régulation et amélioration, éventuellement reprogrammation de certaine partie du module*
- f. *Mise au point d'une version finale du processus de labellisation : manuel et procédures*

## **Volet Transfert de technologie/expertise**

### *Ligne d'actions « Typologie d'offres / scénarios d'intervention »*

Validation de la typologie de produits/services auprès des partenaires de terrain voire des futurs clients.

- a. *Essais applicatifs*
- b. *Régulation, amélioration et affinement des méthodes utilisées dans chacun des produits/services*
- c. *Mise au point des fiches produits/services et des supports ad hoc*
- d. *Finalisation des manuels de procédures*

### *Ligne d'actions « Exploration du marché »*

Approfondissement de l'étude de faisabilité pour la création d'une spin-off qui valoriserait des produits et services dans le domaine du testing standardisé.

- a. *Réalisation d'une étude de marché comprenant :*
  - i. *Collecte d'informations sur les marchés ciblés (concurrents déjà présents, budgets consacrés, stratégie de choix de fournisseurs, seuil de prospection...)*
  - ii. *Etude de marché quantitative sur les pratiques d'évaluation en entreprise, les outils utilisés, les fournisseurs et les besoins non encore rencontrés*
- b. *Etude des possibilités de partenariats :*
  - i. *Partenariat avec des organismes de formations*
  - ii. *Partenariat avec des organismes de certification/labellisation*
  - iii. *Partenariat avec des organismes d'audit*

- 
- c. Etude approfondie des aspects contractuels :*
    - i. Contrats de prestation de services
    - ii. Licences d'utilisation d'applications informatiques
    - iii. Conditions de vente
  
  - d. Elaboration de la stratégie marketing :*
    - i. Plan d'affaire sur 3 ans
    - ii. Stratégie de communication :
      - 1. *Création des logos*
      - 2. *Définition des médias de communication*
      - 3. *Création des medias*
      - 4. *Plan de communication*
    - iii. Etude financière :
      - 1. *Définition des « break even point » par couple produit-marché*
      - 2. *Détermination des coûts de maintenance des applications, de sous-traitance, de location, d'exploitation*
  
  - e. Définition des conditions d'exploitation de la nouvelle spin-off*
    - i. Etude de différents scénarios d'implantation
    - ii. Définition de la meilleure structure d'implantation

## 9.2. Calendrier de réalisation (une page)

		Recherche & Développement		Transfert de technologie / d'expertise		
Mois		[1] Création et validation d'une méthode d'audit de systèmes de testing à utiliser en amont d'e-C&QCST	[2] Création et validation d'un module de labellisation à utiliser en aval de la plateforme e-C&QCST	[3] Création et validation d'une typologie d'offres de services / de scénarios d'interventions pour la future spin-off	[4] Exploration du marché et étude de faisabilité liée à la création de la future spin-off	
Phase exploratoire	1	Formalisation de normes, définition des cadres d'application, identification des indicateurs des niveaux d'adéquation des normes aux contextes	Conception grille de labellisation liée aux modèles SMART	Formaliser des lignes de produits et de services orientés vers l'évaluation en support de la plateforme, du module de labellisation, d'autres applications SMART	Réaliser une étude de marché sur les services et outils d'évaluation	
	2					
	3		Cahier des charges module informatisé de labellisation		Définir les contenus exacts et les champs d'application de ces produits et services et leurs supports <i>ad hoc</i>	Réaliser une étude de marché sur les normes, les agrégations, les certifications et labels
	4					
	5					
	6	Définition des procédures permettant de vérifier les niveaux d'adéquation, développement des supports pour la méthode d'audit	Programmation du module informatisé de labellisation relié à la plateforme e-C&QCST	Définir un avant projet de stratégies marketing		
	7					
	8					
	9					
	10					
	11					
	12					
Phase opérationnelle	13	Expérimentation des procédures de l'avant projet du manuel d'audit auprès des futurs clients potentiels, évaluation de la faisabilité de l'implémentation effective sur le terrain	Présentation de la méthode de labellisation auprès de futurs clients de la spin-off, labellisation expérimentale de tests, essai du module informatisé de labellisation dans le cadre de cette expérimentation	Validation des produits et services via des essais applicatifs chez des clients potentiels	Etude de marché <i>field research</i> , étude financière (coût de location, sous-traitance...), établissement de la stratégie commerciale, étude des aspects contractuels (contrats type, conditions de vente...), étude des réglementations en matière d'opérateurs de formation, en matière de certification et de labellisation	
	14					
	15					
	16					
	17					
	18	Régulation et amélioration de la méthode d'audit (normes, procédures, indicateurs, supports...), création du manuel d'audit dans sa version finale	Amélioration du processus et du programme informatisé de labellisation	Amélioration et affinements des produits/services	Etude de la localisation et de la structure nécessaire	
	19					
	20					
	21					
	22					
	23					
	24					