

CHAPITRE 2

La plateforme ExAMS, un assessment management system pour instrumenter la construction et la gestion de la qualité des évaluations des apprentissages

JEAN-LUC GILLES

Haute école pédagogique du canton de Vaud, Lausanne, Suisse

PASCAL DETROZ – VINCIANNE CRAHAY

Université de Liège, Belgique

SALVATORE TINNIRELLO

Assessment Systems and e-Solutions Group, Liège, Belgique

PIERRE BONNET

Université de Liège, Belgique

Les préoccupations relatives à la qualité de l'évaluation des apprentissages que nous vivons actuellement peuvent être resituées dans le contexte d'un mouvement mondial d'introduction des principes de gestion de la qualité qui pénètre les activités liées à l'éducation et à la formation depuis les années 1990 (Ramsden 1991; Nightingale et O'Neil 1994; Kaufman et collab. 1995; Zink et Schmidt 1995; Segers et collab. 2003). Les pratiques évaluatives des établissements universitaires n'échappent évidemment pas à cette tendance récente; c'est ce que soulignaient Blais et ses collègues en 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.

Par ailleurs, la recherche montre que les enseignants et les apprenants apprécient les instruments d'évaluation des apprentissages lorsqu'ils sont valides, fidèles, sensibles, diagnostiques, équitables, praticables et authentiques (Kirkland 1971; Kulik et Kulik 1987; Leclercq et Gilles 1993; Gilles et Leclercq 1995; Castaigne et collab. 2001; Gilles 2002; De Souza et Fleming 2003; Phelps 2005; Jönsson et collab. 2007a, 2007b; Draaijer et Hartog 2007; Gilles 2010). De toute évidence, la prise en compte de ces qualités docimologiques et de leur praticabilité peut être facilitée par l'utilisation de procédures informatisées. Cependant, la qualité de l'évaluation des apprentissages ne peut être pleinement assurée que si ces procédures informatisées s'insèrent dans un processus structuré en plusieurs phases de mise au point (Gilles 2002; Gilles et collab. 2004; Gilles et collab. 2005; Downing et Haladyna 2006; Crahay et Gilles 2007; Gilles et Detroz 2008; Gilles 2010). Nous décrivons dans la section suivante comment nous envisageons ce processus en fonction de huit étapes qui s'insèrent dans un cycle « en spirale qualité ». Nous utilisons le terme « spirale » dans la mesure où la dernière étape consiste à faire le point sur les différents aspects du dispositif d'évaluation qui pourraient être améliorés. Une perspective qui, d'épreuve en épreuve, place l'évaluateur dans une posture réflexive qui l'amène à prendre du recul vis-à-vis de l'instrument d'évaluation et de son processus de construction, en vue d'augmenter la qualité des prochains cycles de productions d'épreuves.

Dans le cas de la plateforme ExAMS que nous décrivons dans la troisième section du présent chapitre, l'informatisation du cycle de construction des outils d'évaluation des apprentissages a pu être menée à bien grâce au soutien de la Région Wallonne de Belgique qui depuis 2004 a subventionné trois projets de « Recherche et Développement » successifs. Le premier projet intitulé *Electronic Construction and Quality Control in Standardized Testing* a été confié à l'unité de soutien logistique et de recherche du Système Méthodologique d'Aide à la Réalisation de Tests (SMART)¹ de l'Université de Liège (ULg) en partenariat avec la cellule de Gestion des Compétences de HEC – École de Gestion ULg. En trois ans, soit de 2004 à 2007, les travaux de l'équipe ont permis d'aboutir à une première version d'une plateforme soit. un logiciel libre (*open source*) utilisable par l'entremise du Web et intitulée ExAMS 1.0 (*Examination Assessment Management System* – version 1.0). Le deuxième projet intitulé « Logiciel Innovant » s'est déroulé de 2008 à 2009 et a été confié par la Région Wallonne à la *spin off* ASSESS

1. Le SMART possède une longue expérience dans le domaine de l'évaluation des acquis des apprenants (concours de recrutement, examens, tests de prérequis, post-test...) et de l'évaluation de la qualité des enseignements. L'équipe apporte un soutien méthodologique et logistique spécifique à l'Université de Liège (www.smart.ulg.ac.be).

GROUP S.A.² en partenariat avec le SMART. Ce projet a permis, d'une part d'informatiser les procédures docimologiques afin de gérer de façon semi-automatique les Questions à Réponses Ouvertes Longues (QROL) et, d'autre part, d'améliorer l'interopérabilité de la plateforme ExAMS. Enfin, quant au troisième projet, il est intitulé *Quality Design For Competencies Assessment (QDFCA)* et a été confié à ASSESS GROUP S.A. en partenariat avec le SMART et le Centre d'Excellence en Technologie de l'Information et de la Communication (CETIC)³. Ce projet Recherche et Développement a débuté en 2009 et se terminera en 2012. L'objectif est d'implémenter dans le cycle de construction et de gestion de la qualité des procédures docimologiques en conformité avec les principes de l'évaluation dans une approche par compétences. Nous reviendrons à la fin du présent chapitre sur les enjeux liés à l'introduction dans ExAMS des principes de l'évaluation dans une approche par compétences.

LE CYCLE DE CONSTRUCTION ET DE GESTION DE QUALITÉ DE L'ÉVALUATION DES APPRENTISSAGES

Nous situons la production d'une démarche d'évaluation des apprentissages dans le cadre d'un paradigme qui inclut notre conception d'une approche de qualité en docimologie basée sur quatre facteurs clés à prendre en compte dans la mise au point des épreuves. En toile de fond de ces facteurs clés, nous proposons également de tenir compte des cadres contextuels et axiologiques de l'évaluation. Ces éléments forment les fondements de la démarche de qualité que nous allons décrire dans la présente section. C'est au sein du facteur clé « modèles théoriques » que s'insère le cycle de construction et de gestion qualité de l'évaluation des apprentissages (Gilles 2002; Gilles et collab. 2004).

Le paradigme de l'approche qualité

Les acteurs engagés dans une démarche d'évaluation (décideurs, spécialistes en évaluation, experts dans un domaine donné, enseignants ou

-
2. Créée en 2006, la société Assess Group S.A. (*Assessment Systems and e-Solutions Group*) est une entreprise par essaimage (*spin-off*) issue de l'Université de Liège. Elle a été créée par trois membres appartenant à l'époque à l'unité de soutien logistique et de recherche SMART (Jean-Luc Gilles [promoteur du projet], Salvatore Tinnirello et Pascal Detroz). Assess Group S.A. propose ses services dans le domaine de l'évaluation au monde des entreprises et des administrations (www.assess-group.be).
 3. Le CETIC est actif en recherche appliquée en génie logiciel, en technologies orientées service et en systèmes embarqués. Il est un agent de connexion, de transfert de technologie entre recherche universitaire et entreprises (www.cetic.be).

formateurs, évalués...) évoluent dans des contextes particuliers qui influencent les décisions qu'ils prennent. L'évaluation n'est pas neutre, elle s'insère dans des systèmes scolaires ou de formation qui lui assignent des fonctions (par exemple pronostique, formative, sommative...) et qui utilisent les informations qu'elle génère pour alimenter les prises de décisions. En toile de fond, les valeurs, au sens axiologique du terme, influencent ces systèmes et par conséquent l'usage qui est fait de l'évaluation et de ses conclusions. Par exemple, dans un système de formation au sein d'une entreprise, un responsable des ressources humaines amené à mettre au point un test en vue de sélectionner de nouveaux employés devra tenir compte des caractéristiques d'une série de variables contextuelles : administratives, socioéconomiques, socioculturelles, etc. Elles-mêmes balisées par une série de cadres axiologiques déterminant des valeurs véhiculées au sein de l'entreprise. Dans le cas d'une démarche d'évaluation se déroulant dans un établissement d'enseignement, les variables contextuelles pourraient être très différentes d'un établissement à l'autre, les cadres axiologiques aussi. Imaginons deux établissements d'enseignement universitaire l'un défendant une vision socioconstructiviste des apprentissages en en faisant un label pour se démarquer, et l'autre prônant une vision plus traditionnelle. Les épreuves, pour peu que leurs concepteurs maintiennent une certaine cohérence entre les aspects didactiques et didactologiques, prendraient des allures très différentes.

Quel que soit le secteur (établissements d'enseignement, entreprises, administrations...) et quels que soient les cadres axiologiques et les variables contextuelles, quatre facteurs clés détermineront l'efficacité du processus de mise au point d'un outil pour l'évaluation :

1. la prise en compte des besoins des acteurs ;
2. le degré d'adoption de standards et de pratiques efficaces validés par la recherche
3. une gestion efficiente des ressources humaines et matérielles disponibles ;
4. l'utilisation de modèles théoriques adaptés à la situation.

La figure 1 ci-dessous présente une vue schématisée de l'approche qualité proposée (Gilles 2002). Les traits unissant ces quatre facteurs clés montrent l'importance que revêt leur prise en compte simultanée et cohérente. Les besoins des acteurs sont placés au milieu du schéma, car ils se situent au centre de nos préoccupations en vue d'assurer le succès des opérations liées à l'évaluation.

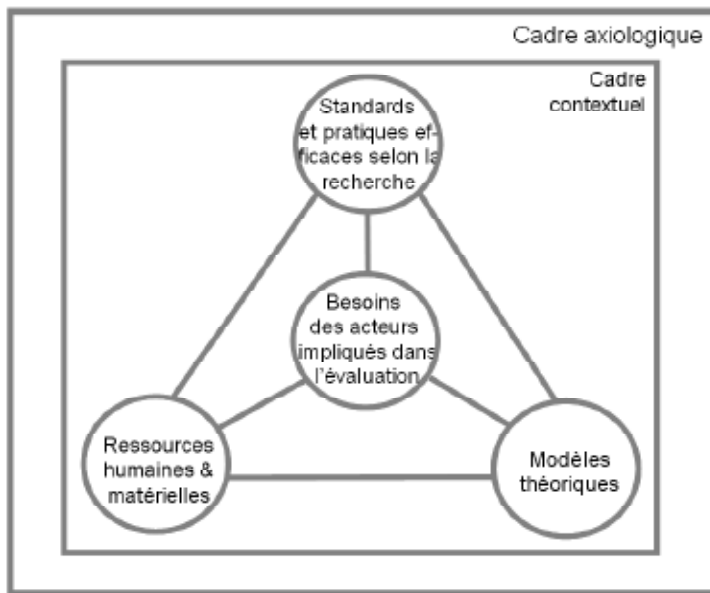


FIGURE 1 – Schéma de l'approche qualité (Gilles 2002, 2010)

Le facteur situé au sommet concerne le respect de critères de qualité docimologique. De ce point de vue, les *Standards of Educational and Psychological Testing* (American Educational Research Association, American Psychological Association et National Council on Measurement in Education 1999) constituent par exemple une excellente référence. Les pratiques efficaces validées par la recherche en docimologie ont également fait l'objet de recensements et de manuels utiles à tous ceux qui sont engagés dans l'élaboration d'outils d'évaluation (Downing et Haladyna 2006; Nitko et Brookhart 2007). Les compétences d'un expert en évaluation des acquis reposent ainsi en grande partie sur sa capacité à mobiliser à bon escient des savoirs et des savoir-faire en rapport avec des critères de qualité reconnus.

En ce qui concerne les ressources humaines nécessaires à la démarche d'évaluation des apprentissages, nous distinguons cinq rôles fondamentaux: coordination, expertise docimologique, expertise du domaine, soutien logistique et public évalué. Nous verrons plus loin que ces rôles peuvent être distribués au sein de la plateforme ExAMS, ce qui permet et facilite alors le travail collaboratif. Notons que dans certains contextes, les quatre premiers rôles peuvent être tenus par une seule et même personne. Par exemple, dans le cas d'un enseignant amené à concevoir et à proposer une épreuve au sein de sa classe. Dans d'autres cas, par exemple dans le contexte d'une entreprise, la coordination pourrait être prise en charge par le responsable des ressources

humaines, l'expertise docimologique par un consultant externe, l'expertise du domaine par un technicien au sein de l'entreprise et enfin le soutien logistique par le personnel du département de reprographie. À maintes reprises, nous avons pu observer ce type de distribution des rôles dans le cadre d'accompagnements docimologiques offerts par ASSESS GROUP S.A. souvent amenée à fournir l'expertise docimologique ou le soutien logistique pour des entreprises tierces.

En ce qui concerne les ressources matérielles, l'informatique offre une grande variété de possibilités d'automatisation, de semi-automatisation ou de soutien efficace dans la réalisation et la passation d'épreuves et de tests. La figure 2 ci-dessous montre la catégorisation des méthodes d'administration de tests proposée par Thompson (2008).

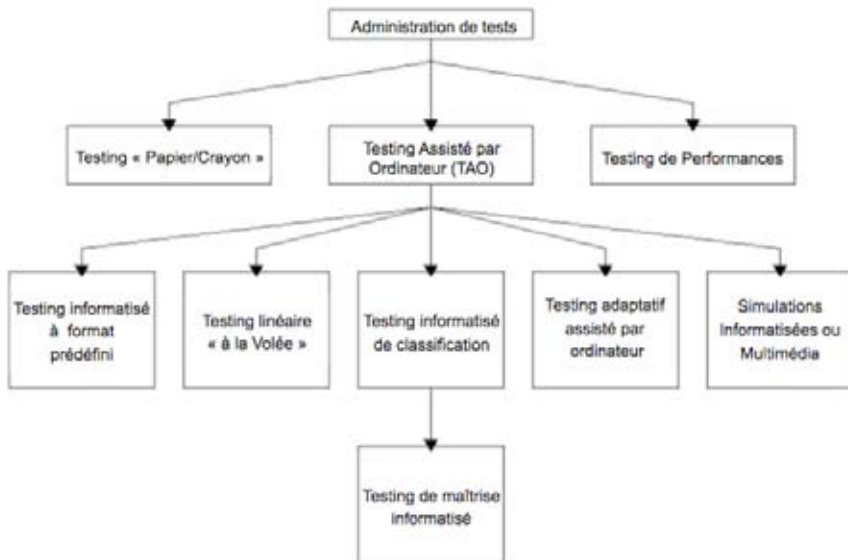


FIGURE 2 – Catégorisation des méthodes d'administration de test.
Traduction libre des catégories et sous-catégories
proposées par Thompson (2008).

La plateforme ExAMS procure une aide efficace pour la création de tests classiques, qu'ils soient administrés sous forme « papier/crayon » ou « clavier/écran ». Ce qui dans le deuxième cas (catégorie testing assisté par ordinateur [TAO]) offre alors la possibilité d'une notation automatisée et de rétroactions immédiates (sous-catégorie testing informatisé à format prédéfini [TIFP]). Les modèles de notation et d'analyse utilisés dans ExAMS reposent sur la

théorie classique des tests (TCT), un des deux paradigmes dominant en psychométrie avec la théorie des réponses aux items (TRI).

Les autres catégories du schéma de Thompson, testing informatisé de classification (*Computerized Classification Testing – CCT*) et testing adaptatif assisté par ordinateur (*Computerized Adaptive Testing – CAT*), fournissent l’occasion, lorsque les échantillons sont suffisants, d’utiliser les algorithmes de notation du modèle de Rasch (1953, 1966). Ces modalités de *testing*⁴ ne sont pas encore disponibles dans ExAMS. Tout comme la catégorie testing linéaire « à la volée » (*Linear On the Fly Testing – LOFT*) (Stocking et collab. 2000) parfois reprise sous l’appellation « *Automated Test Assembly* » (ATA) (Chuan-Ju Lin 2008; Cor et collab. 2009, Riopel, Raïche, Pilote et Potvin 2009). Le testing linéaire « à la volée » repose sur des banques d’items consécutives à partir desquelles le système construit le test en direct au moment de la passation, à la volée, en fonction de caractéristiques propres à chaque candidat (comme les sous-catégories *testing adaptatif assisté par ordinateur* et *testing informatisé de classification*), ce qui aboutit à des tests différents selon les candidats.

L’informatisation des procédures d’évaluation de performances complexes, la catégorie *Testing de performances* sur le schéma de Thompson, est au cœur du projet *Quality Design For Competencies Assessment* (QDFCA). Un de nos objectifs de recherche dans le cadre de QDFCA consiste à créer des liens entre les deux catégories *testing de performances* et *simulations informatisées ou multimédia* à travers de nouvelles fonctionnalités à introduire dans ExAMS en vue de faciliter la gestion, la correction et la notation de performances complexes qui, par exemple, auront préalablement été mises en format vidéo. Ces nouvelles fonctionnalités participeront à l’instrumentation de l’évaluation dans une approche par compétences.

Le cycle en huit étapes de construction des outils d’évaluation

Le quatrième facteur clé de l’approche qualité reprend les modèles théoriques qui doivent guider nos actions lors de la construction d’une épreuve, et ce, en rapport avec les trois autres facteurs clés et en tenant compte des cadres axiologique et contextuel. Dans cette optique, en ingénierie docimologique, nous avons mis au point une approche « processus » pour élaborer des outils d’évaluation de qualité (Gilles et Leclercq 1995; Gilles 2002; Gilles et Lovinfosse 2004; Gilles et collab. 2004 2005; Crahay et collab. 2006; Gilles et collab. 2008a, 2008b). D’autres docimologues

4. Pour plus d’informations à propos des différentes catégories, nous renvoyons à l’article de Thomson (2008)

proposent des approches du même type, par exemple Downing et Haladyna (1997, 2006), mais à la différence de ces derniers qui proposent 12 étapes de construction, notre modèle structure le processus de réalisation de l'évaluation en huit étapes qui sont implémentées dans la plateforme ExAMS.

1) *L'étape de l'analyse*

Cette étape impose l'identification des compétences à évaluer, de leurs caractéristiques en termes de savoirs, de savoir-faire et de ressources qui devront être mobilisés par le candidat, ainsi que de l'importance de ces aspects dans l'épreuve à construire. Elle implique par ailleurs l'analyse des composantes des situations dans lesquelles le candidat sera amené à performer. À l'issue de cette étape, la plateforme ExAMS produit une table de spécification croisant les points de contenus avec les processus mentaux mobilisés et les modalités de questionnement sélectionnées. L'idée est de s'assurer que la démarche couvrira les aspects importants que l'évaluateur s'est donné pour objectif d'évaluer. C'est toute la problématique de la validité au sens où elle est définie dans les *Standards for educational and psychological testing* (AERA, APA et NCME 1999, p. 9-24) qui est en jeu à cette étape.

2) *L'étape du design*

Elle intègre les résultats de l'étape d'analyse et amène l'évaluateur à choisir les modalités de questionnement qu'il mettra en œuvre dans son épreuve. Son dispositif peut mobiliser diverses techniques qu'il peut être amené à combiner et dont l'utilisation sera planifiée dans le temps: questions de production impliquant des échelles descriptives, méthodologie du projet, portfolio, observations, tests standardisés traditionnels, analyses de cas, etc. La réflexion portera également sur les principes de correction qui seront appliqués pour assurer une bonne fidélité des outils de jugement des performances. On trouve ici des préoccupations que Downing et Haladyna (2006) rangent dans leur étape 3, *Test specification*, et que l'on trouve aussi dans les *Standards for educational and psychological testing* sous le vocable *Development and evaluation of the test specification* (AERA, APA et NCME 1999, p. 37-38).

3) *L'étape de l'élaboration des items*

Cette étape doit permettre d'alimenter la démarche d'évaluation en modalités en vue de solliciter et de recueillir les performances des candidats en fonction des décisions prises aux deux étapes précédentes. Par exemple, dans le contexte d'une épreuve, dans une approche par compétences, l'évaluateur pourrait être amené à alimenter le dispositif d'évaluation en questions

ouvertes en rapport avec des tâches complexes définies sur le plan de leurs caractéristiques générales à l'étape 2 et portant sur des compétences et des situations proches de la réalité identifiées à l'étape 1. Dans le cadre de la construction de questions à réponses ouvertes longues (QROL) à cette étape d'élaboration des items, des échelles descriptives de cotation avec leurs critères d'appréciation et leurs indicateurs de performances devront également être explicitées. C'est toute la problématique de la création, de la gestion et du contrôle de la qualité *a priori* des items qui est concernée par cette étape où les collaborations sont fréquentes entre docimologues et spécialistes du domaine couvert par l'épreuve (AERA, APA et NCME 1999, pp. 38-39). Notre troisième étape, celle de l'élaboration des items correspond à l'étape 4, *Item development*, chez Downing et Haladyna (2006, p. 10-12).

4) *L'étape de la préparation*

Dans le cas où l'évaluateur juge cette étape pertinente, il offrira aux candidats des informations relatives à l'épreuve ainsi que des possibilités de « performer à blanc ». Il s'agit ici d'offrir aux candidats des possibilités de se familiariser aux types de tâches proposées mais aussi d'offrir aux évaluateurs des occasions d'expérimenter les procédures dans le cadre d'éventuels entraînements. Depuis des décennies, cette étape est préconisée par les docimologues lorsque les procédures sont sophistiquées ou les évalués peu habitués aux méthodes proposées. Par exemple, Shufford et collab. (1966) préconisent systématiquement cette étape, entre autres recommandations méthodologiques, dans le cas d'épreuves où les candidats sont invités à accompagner chaque réponse d'un pourcentage de certitude (Leclercq et Gilles 2001).

5) *L'étape de la passation*

C'est lors de cette cinquième étape du processus que l'évaluateur sera amené à assembler les questions de l'épreuve sélectionnée dans la banque de questions. Il effectuera également les derniers paramétrages du dispositif et diffusera les épreuves aux candidats. La plateforme ExAMS permet de gérer les passations en ligne (par exemple dans un local surveillé où les candidats ne peuvent avoir accès qu'à la plateforme) ainsi que toute la logistique liée aux administrations sous forme papier-crayon (par exemple dans le cadre d'un examen ou d'un concours regroupant plusieurs centaines de personnes au même moment dans un même lieu). En ce qui concerne les dispositifs en ligne, la plateforme évolue aujourd'hui grâce au projet QDFCA vers des interfaces de passation électroniques facilitant une approche par compétences et offrant des possibilités de médiatisation (images, vidéo, son) simulant la réalité et permettant d'interagir avec ces situations médiatisées. L'idée est de

permettre, à terme, la prise en charge dans ExAMS de tâches de la catégorie *Computerized Simulation or Multimedia* proposée par Thompson (2008) dans sa classification des méthodes d'administration de tests.

6) *L'étape de la correction*

Cette étape amènera les évaluateurs à juger des performances des candidats à la suite de la passation. Dans le cadre de performances complexes, c'est toute la problématique des interférences liées aux interactions « candidat x juge x situation » qui devra être gérée, instrumentée et maîtrisée. La recherche « Logiciel innovant » et les fonctionnalités QROL avec la gestion des échelles descriptives qu'elle introduit dans la plateforme permettent aujourd'hui au responsable qui supervise l'utilisation d'ExAMS de piloter la qualité de la concordance interjuges et intrajuges. Tout en facilitant le travail collaboratif dans la correction des productions, ExAMS offre des fonctionnalités de vérification de la qualité du travail de correction en temps réel et, le cas échéant, permet au responsable de la démarche d'avertir les correcteurs des problèmes qu'il voit apparaître sur le plan du monitoring de la correction. Les mesures d'ajustement qui peuvent ainsi être prises lors du suivi de la correction, contribuent à réduire les problèmes de fidélité qui ont toujours préoccupé les docimologues et qui font l'objet d'un ensemble de recommandations dans les *Standards for educational and psychological testing* (AERA, APA et NCME 1999, p. 25-36). Lors de cette étape de correction, ExAMS propose des contrôles qualité *a posteriori* des résultats à l'aide d'indices classiques (indices de cohérence interne). Des rectifications peuvent ensuite être mises en œuvre de manière à améliorer la fidélité.

7) *L'étape de la rétroaction*

Après qu'une version définitive des résultats a été élaborée à l'étape précédente, les candidats recevront des informations sur leurs performances à l'épreuve. Si celle-ci s'insère dans un processus d'évaluation formative, l'évaluateur devra veiller à offrir des rétroactions diagnostiques, personnalisées et détaillées en vue de faciliter le bilan des forces et des faiblesses de l'apprenant. Dans le contexte d'ExAMS, ce travail est grandement facilité par la table de spécifications élaborée lors de la première étape, celle de l'« Analyse » et qui permet d'identifier ici les items qui relèvent des différentes catégories de contenus visés et de processus mentaux mobilisés. À partir d'un repérage automatisé des performances des candidats à ces différentes catégories de contenus et de processus mentaux le responsable de la démarche peut offrir une rétroaction diagnostique très détaillée aux candidats. La rapidité de l'envoi des rétroactions est également cruciale et, de ce point de vue, les possibilités

offertes par ExAMS permettent des délais très courts entre la passation de l'épreuve et l'envoi de rétroactions personnalisées aux candidats.

8) *L'étape de la macro régulation*

Cette huitième et dernière étape est cruciale pour les personnes engagées dans le processus de création de l'épreuve. L'idée est de prendre un temps de recul réflexif en fin de processus pour effectuer une analyse systématique de ce qui a été réalisé lors des sept étapes précédentes. Les difficultés éprouvées seront discutées et des pistes de régulations proposées en vue d'améliorer la construction des prochaines épreuves. C'est en cela que nous parlons d'un cycle en huit étapes de mise au point des outils d'évaluation, la dernière étape permettant de réguler le travail réalisé aux étapes précédentes. L'enjeu n'est pas des moindres, il s'agit d'amener les évaluateurs à s'engager dans une démarche continue de remise en question et de recherche d'améliorations de la qualité docimologique des épreuves qu'ils construisent.

L'INSTRUMENTATION DU MODÈLE DE CONSTRUCTION ET DE GESTION DE LA QUALITÉ DES APPRENTISSAGES À L'AIDE DE LA PLATEFORME *OPEN SOURCE* EXAMS

La plateforme ExAMS a pour objectif d'informatiser le cycle en huit étapes, décrit dans la section précédente, en vue de faciliter la construction et la gestion d'outils de qualité et de permettre aux candidats d'en bénéficier. Au travers des procédures informatisées et des contrôles de la qualité présents à chacune des étapes, la plateforme accompagne le responsable de la démarche depuis l'analyse initiale des compétences à évaluer jusqu'à l'envoi des rétroactions aux candidats, en passant par l'élaboration de la forme de l'épreuve, la création des questions, la passation des épreuves et la correction. Elle lui permet d'agir de manière autonome dans la gestion de la démarche tout en favorisant un travail collaboratif à distance s'il souhaite s'entourer d'une équipe de personnes. Les candidats, quant à eux, sont en mesure de recevoir de l'information au sujet de leur évaluation, de passer des tests d'entraînement ou des tests certificatifs ainsi que de consulter des rétroactions diagnostiques, précises et rapides sur leurs performances.

Utilité d'ExAMS pour le responsable de la démarche d'évaluation

La plateforme ExAMS couvre de vastes besoins en évaluation. Par exemple, elle sera utile à un enseignant qui cherche à vérifier les connaissances de ses étudiants pour une matière donnée. Elle sera également performante en entreprise pour obtenir des informations sur les apprentissages cognitifs

après une formation ou encore dans le cadre d'un dépistage des compétences des travailleurs dans un contexte de gestion des connaissances (*knowledge management*). Dans tous les cas, elle offrira une démarche assurant la qualité des scores obtenus et des rétroactions offertes.

Pour présenter ExAMS et illustrer ses fonctionnalités, nous partirons d'un exemple concret concernant un professeur d'anatomie donnant cours aux premiers bacheliers et soucieux de les évaluer avec un haut niveau de qualité. Dans notre exemple, tiré d'un cas réel, les étudiants sont au nombre de 600 et deux assistants aident le professeur dans sa tâche. Ce professeur, conformément au cycle de construction et de gestion des outils d'évaluation qui a été présenté dans la section précédente, va commencer par mener une analyse de la matière qu'il a enseignée et sur laquelle va porter la collecte des données. Il commence par établir la liste des points qu'il voudrait évaluer [PE] (en utilisant par exemple la table des matières détaillées de son cours) et leur accorde ensuite une priorité. Puis, il détermine les catégories de performance [CP] qu'il vise chez les étudiants (en s'inspirant par exemple de la taxonomie de Bloom) et relie enfin chaque [PE] à au moins une [CP]. Bien qu'il ait pu confier ce travail à l'un de ses assistants, il décide, en tant que responsable du cours, d'effectuer ce travail seul. Il se dote ainsi d'un tableau appelé « table de spécifications », lui permettant de cibler ce qu'il vise et de contrôler la validité de sa démarche.

Points	Connaissances	Compréhension	Application
1. La position anatomique de référence [1]	✓		
2. Les vues anatomiques [2.1] [2.2]	✓	✓	
3. Les 3 axes de référence [3]	✓		
4. Les 3 plans de référence [4.1] [4.2]	✓	✓	
5. Les 2 types de plans obliques [5.1] [5.2] [5.3]	✓	✓	
6 a. Section anatomique [6.1]	✓		
6 b. Franche de section [6.2]	✓		
6 c. Coupe anatomique [6.3]	✓		
6 d. Distinction entre tranche de section et coupe anatomique [6.4]		✓	
7. Règles d'orientation des coupes [7.1] [7.2]	✓	✓	
8 a. Définition de la notion des Coupes sérées [8.1]	✓		
8 b. Le volume (géométrique) d'une structure anatomique sur base de l'étude de coupes sérées [8.2] [8.3]	✓		✓
9 a. Utilisation des coupes (descripteur, représentation) [9.1] [9.2]		✓	✓
9 b. Série de coupes doivent une structure anatomique donnée [9.3] [9.4]		✓	✓
9 c. Structure anatomique sur base d'une série de coupes donnée [9.5] [9.6]		✓	✓
10. La nomenclature anatomique internationale [10]	✓		
11 a. L'abréviation [11.1] [11.3]	✓		

FIGURE 3 – Extrait d'une table de spécifications présentée dans ExAMS.

Chaque coche dans le tableau présente un croisement entre un point à évaluer [PE] et une catégorie de performance [CP] pouvant potentiellement faire l'objet d'une question

C'est ensuite à l'étape Design que des modalités de questionnement [MQ] vont être définies et intégrées dans la table de spécification. Celles-ci sont constituées par des formats de questionnement tels que QCM, vrai faux, Questions à réponses ouvertes longues (QROL) parfois accompagnés d'options telles que: le recours à des Solutions générales implicites (SGI) dans le cas de QCM; la possibilité d'organiser un examen à Livre ouvert (LO); l'utilisation de la technique des Degrés de certitude (DC), etc.

Ces modalités de questionnement [MQ] seront définies en cohérence avec les éléments de la table de spécification. Par exemple, si l'enseignant constate à l'aide de sa table de spécification, qu'il veut essentiellement évaluer des connaissances chez ses étudiants, il choisira comme modalité de questionnement les QCM. Ceux-ci étant tout à fait indiqués pour évaluer sa matière avec un bon niveau de couverture. En revanche, si sa table de spécifications indique qu'il est plus soucieux d'évaluer des performances complexes, par exemple la capacité des étudiants à synthétiser de l'information et à produire des raisonnements, il optera pour des QROL. Le plus souvent, l'enseignant visera des niveaux de performances divers et utilisera plusieurs modalités de questionnement qu'il prendra le soin de définir. Nous préconisons de telles approches combinées lorsque leur pertinence est démontrée lors des premières étapes du cycle dans un contexte de construction des examens « à l'aide d'un processus méthodique dont les régulations permettent une amélioration continue des épreuves » (Gilles 2002, p. 35).

Le responsable associera ensuite chaque modalité de questionnement aux deux premiers éléments [PE] x [CP] de la table de spécification. Il obtiendra alors des trinômes [PE] x [CP] x [MQ] qui orienteront la création des questions à l'étape suivante.

The screenshot shows the ExAMS software interface. At the top, there is a navigation menu with options: Analyse, Design, Questions, Information, Test, Correction, Feedback, Régulation. Below this, there are search filters for 'Text', 'VF', and 'ID'. The main part of the image is a table titled 'C1_D Ana-' with columns for 'Connaissance', 'Compréhension', and 'Application'. Each column has sub-columns for 'Score attribué à l'élève' and 'QCM SCQ OC'. The table lists 10 items related to anatomy, such as 'La position anatomique de référence [1]', 'Les vues anatomiques [2.1] [2.2]', etc., with green dots indicating the associated assessment types.

C1_D Ana-	Point	Connaissance		Compréhension		Application	
		Score attribué à l'élève	QCM SCQ OC	Score attribué à l'élève	QCM SCQ OC	Score attribué à l'élève	QCM SCQ OC
1. La position anatomique de référence [1]	1	●	●				
2. Les vues anatomiques [2.1] [2.2]	1	●	●	●		●	
3. Les 3 axes de référence [3]	1	●	●				
4. Les 3 plans de référence [4.1] [4.2]	1	●	●				
5. Les 2 Tables de plans adossés [5-1] [5-2] [5-3]	1	●	●			●	
6.A. Section anatomique [6.1]	1	●	●				
6.B. Tranche de section [6.2]	1	●	●				
6.C. Coupe anatomique [6.3]	1	●	●				
6.D. Différence entre tranche de section et coupe anatomique [6.4]	1			●			●
7. Règles d'orientation des coupes [7.1] [7.2]	1	●	●				
8.A. Définition de la notion des Coupes adossés [8-1]	1	●	●				

FIGURE 4 – Extrait d'une table de spécifications présentée dans ExAMS. Chaque point dans le tableau présente un trinôme [PE] x [CP] x [MQ] et permet d'accéder à l'interface de construction du type de questionnement qui lui correspond

L'enseignant possède alors toutes les données nécessaires pour créer les questions de son épreuve. Après une analyse rigoureuse, il a identifié les éléments de la matière qu'il souhaite évaluer, les catégories de processus mentaux qu'il cherche à évaluer chez ses étudiants ainsi que les modalités de questionnement adéquates. Il est prêt pour alimenter sa banque de questions en nouveaux items ou pour aller y puiser les questions qui existent déjà et qu'il souhaite réutiliser.

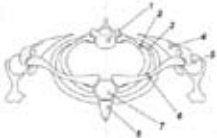
Imaginons un cas de figure où le responsable de l'évaluation sollicite l'aide de deux assistants. Il leur demande de créer ensemble, pour chacun des trinômes de la table de spécifications, plusieurs questions en vue d'alimenter la banque de questions. Il délègue cette tâche sans crainte sachant que ce travail de création est balisé par la table de spécification et que les questions devront d'abord être validées par lui avant de pouvoir être utilisées dans un test. Dans l'interface prévue pour les QCM, les assistants devront encoder l'amorce de la question et les différentes propositions en signalant la réponse correcte et les rétroactions éventuelles associées à chaque solution. Des médias pourront être associés à chacun de ces éléments. Par exemple des images et, sous certaines conditions, du son ou de la vidéo. Dans le contexte d'ExAMS, une QCM est donc définie comme un objet comportant

une amorce et des propositions de réponses parmi lesquelles figure la solution correcte, le tout se caractérisant par un ou plusieurs trinômes de référence [PE] x [CP] x [MQ]. Ainsi, chaque question est clairement associée à une partie de la table de spécifications.

Afficher la question

CS-1

Sur la figure ci-dessous, on peut affirmer que




1. ✓ Les "6" s'étend de "7" à "4" suivant une direction oblique en dehors, en arrière et en haut.
2. Les "3" correspondent à T4.
3. qu'il est possible d'identifier les os "5" gauche et droit.
4. le muscle "7" est exclusivement adducteur de "3".
5. Solution Générale Implite 6.
6. Solution Générale Implite 7.

FIGURE 5 – Visualisation d'un extrait d'une Question à choix multiple

Dans l'interface prévue pour les Questions à réponses ouvertes longues (QROL), ils pourront encoder la question et spécifier une zone de réponse à l'écran ou sur papier (ExAMS permet des passations sous forme papier-crayon, nous y reviendrons). Ici aussi, ils pourront associer un média à la question. Ils pourront également prévoir une image de fond pour la zone de réponse, dans la perspective, par exemple, d'annoter un schéma.

Consulter la consigne

Introduction : Soit une série de coupes (Figure 1) présentées, dans l'ordre, de A à F et toutes orientées de façon similaire comme indiqué sur la coupe A.



Question : Représentez par une figure et la lettre correspondante (A à F) le plan de coupe utilisé pour obtenir les coupes A à F (6 figures à indiquer, chacune identifiée par une lettre de A à F) (Répondez, en rouge, sur la figure 2-ci)

Critères : Critères visibles avec la consigne sur le questionnaire

Consulter la zone de réponse

Zone de réponse : 

Notes :
Fichier :



Modifier

FIGURE 6 – Extrait d'une Question à réponse ouverte longue (QROL)

La grande différence par rapport aux QCM réside toutefois dans la nécessité d'associer aux QROL des énoncés/critères avec des échelles descriptives de cotation qui formeront la grille de correction de l'épreuve et nécessiteront l'intervention de correcteurs lors de l'appréciation des réponses. Les supports descriptifs dans ExAMS comportent trois niveaux: critères d'appréciation, aspects à apprécier pour chaque critère et indicateurs de performance pour chaque aspect.

Une fois ce travail effectué, le responsable de la démarche va pouvoir valider ou non les questions construites par ses assistants. Pour la validation technique il bénéficiera d'un indice de complétude de la question (vérification automatique de la présence de tous les éléments requis). Il devra également s'assurer de la qualité docimologique des questions et de la pertinence de la grille de correction, éventuellement en ayant recours à l'avis de spécialistes du contenu ou de docimologues. Ces échanges avec des collègues et des professionnels de l'évaluation sont facilités par la plateforme ExAMS qui permet au responsable d'offrir à ses collaborateurs des accès à distance pour certaines fonctionnalités.

Critères	
Titre	Critère 1 : Savoir symboliser un plan de coupe sur un vu de profil à partir de coupes dérivées
Description	
Pondération	1
Aspects	
Titre	Aspect 1.1 : Représentation du plan par une ligne horizontale sur la vue de profil
Description	
Pondération	1
Couverture	9.4. Utilisation des coupes (description, représentation) [9.1] [9.2] » Application » Schéma anatomique à compléter
Indicateurs	
Vertical	0
Oblique 45°	33
Légèrement oblique	56
parfaitement H	100
Aspects	
Titre	Aspect 1.2 : Représentation des différents plans perpendiculaire parallèles
Description	
Pondération	1
Couverture	9.4. Utilisation des coupes (description, représentation) [9.1] [9.2] » Application » Schéma anatomique à compléter
Indicateurs	
Pas de parallélisme	0
Léger défaut de parallélisme	50
Parallélisme parfait	100

FIGURE 7 – Représentation d'un support descriptif liée à une Question à réponse ouverte longue (QROL)

D'un point de vue procédural, ces trois premières étapes d'analyse, de design de l'épreuve et de création des questions, ont amené le responsable à préparer les bases nécessaires à la réalisation de son épreuve. Il est prêt à entrer dans une nouvelle phase plus liée aux candidats potentiels. En effet, il va maintenant s'intéresser à l'information qu'il souhaite fournir à ses étudiants avant la passation de l'épreuve en vue de les y préparer au mieux. Va-t-il leur communiquer la table de spécification, les préparer à l'aide d'un test d'entraînement, leur donner des consignes spécifiques ou encore fournir un document précisant certains aspects des performances attendues lors de l'épreuve? Ces décisions sont de son ressort. Les moments auxquels ces éléments sont mis à la disposition des étudiants peuvent être définis à l'aide d'ExAMS qui informe par courriel de la disponibilité de ces informations.

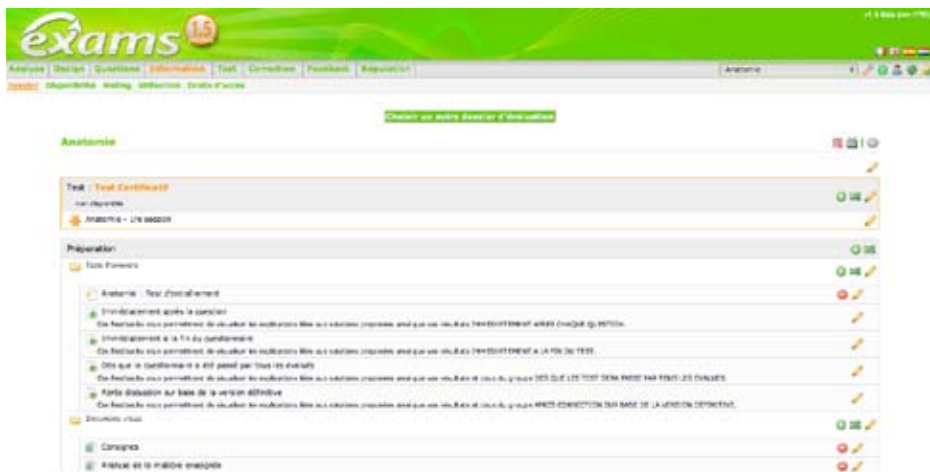


FIGURE 8 – Interface d'ExAMS permettant de gérer les informations et leur disponibilité en vue d'une préparation des évalués avant la passation de l'épreuve

L'étape suivante du cycle de construction permettra au responsable de paramétrer son test. Ce qui signifie qu'il va sélectionner les questions de l'épreuve parmi celles qu'il a validées dans sa banque de questions, qu'il va ensuite les ordonner et enfin choisir un barème de correction. Par exemple, pour un test composé de 20 questions sous forme de QCM accompagnées des degrés de certitude, le barème de correction qui sera utilisé est un tarif « DC » qui permet d'attribuer un score en fonction du degré de certitude associé aux choix de réponse. C'est également lors de cette cinquième étape du cycle qu'il va définir l'option d'une passation en ligne ou sous forme papier-crayon. Dans le cas d'une épreuve en ligne, il devra préciser l'ordre des questions (aléatoire ou fixe, par exemple) et les conditions de mise à disposition du questionnaire pour les étudiants.

Dans le cas d'une épreuve sous forme papier-crayon, il pourra imprimer la version originale du questionnaire automatiquement mis en page par la plateforme ExAMS qui gère également le mélange des questions dans les différentes versions du même questionnaire. Ces fonctionnalités automatisées dans ExAMS offrent un gain de temps considérable aux responsables lors de la phase de préparation des passations. Lors d'une passation électronique, le responsable de la démarche peut se connecter à ExAMS pendant que les candidats répondent aux questions en ligne. L'interface prend alors la forme d'un tableau de bord permettant au responsable d'observer le processus de passation et, le cas échéant, de repérer ou d'anticiper d'éventuels

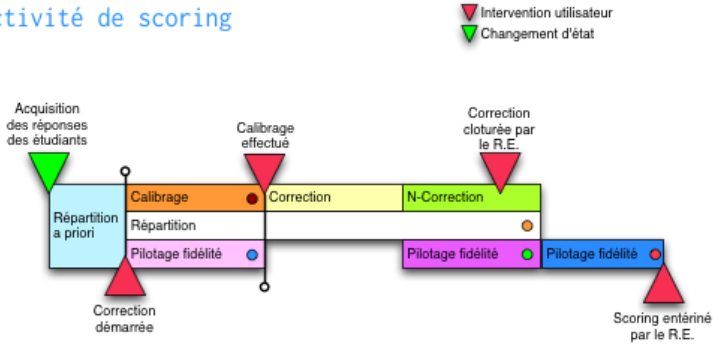
problèmes. Les réponses saisies à l'écran par le candidat sont directement intégrées dans la banque de données et prêtes à être traitées. Dans le cadre d'une passation sous forme papier-crayon, il faudra passer par une phase d'encodage des réponses dans le système. Lorsqu'il s'agit de questionnaires avec des questions à choix multiple, cela peut se faire par l'entremise de l'acquisition automatique des données transcrites par les candidats sur des feuilles spéciales destinées à la lecture optique de marques. Un fichier au format texte comportant les réponses des candidats est alors importé dans la plateforme ExAMS. Dans le cas des QROL il est également possible de numériser les copies des candidats.

Lorsqu'il s'agit de QCM, une fois les réponses introduites dans la base de données celles-ci peuvent être traitées directement et le système offre alors une série de statistiques sur lesquelles nous reviendrons. La procédure de correction est automatisée et consiste à comparer la modalité de réponse choisie par le candidat à celle notifiée comme correcte par le responsable lors de la création de la question. Dans le cas des QROL, il n'est pas possible d'automatiser entièrement la correction (du moins actuellement). L'expertise humaine reste nécessaire pour analyser la production du candidat en fonction des critères élaborés à l'aide des échelles descriptives d'évaluation. Cependant, dans ExAMS la correction des QROL est soutenue par diverses fonctionnalités innovantes. Ainsi, le responsable peut demander à ses assistants de corriger en parallèle l'ensemble des copies et de bénéficier d'une série de statistiques l'informant de la qualité de la cohérence intrajuges et interjuges. Ces informations lui permettent de prendre des mesures en vue de réguler les corrections effectuées par ses assistants et d'en vérifier les effets. Ce pilotage du travail permet d'améliorer la fidélité des corrections.

Cette procédure de correction des QROL mise en place dans ExAMS se subdivise en trois phases. Une première phase de calibrage, durant laquelle les correcteurs vont devoir corriger les mêmes copies, ce qui permettra au responsable de l'évaluation de vérifier et de réguler la concordance des corrections entre les différents correcteurs. Lorsque le niveau de concordance est jugé suffisant, une deuxième phase de production dans le travail de correction peut débuter. La plateforme soumet des copies différentes à chacun des correcteurs dans le but de corriger le plus rapidement possible l'ensemble des copies. Lorsque toutes ont été corrigées au moins une fois, une troisième phase de consolidation des corrections peut commencer au cours de laquelle les copies qui avaient déjà été attribuées à un des correcteurs sont proposées aux autres. Lors de chacune de ces trois phases, le responsable de l'évaluation peut intervenir, tout comme il peut configurer préalablement chacune d'entre elles. Il peut par exemple mener une série d'actions comme supprimer ou ajouter un correcteur, faire en sorte que l'on recorrige des items divergents

(des réponses sur lesquelles les correcteurs ne s'accordent pas), modifier la grille de correction, etc.

QROLs, activité de scoring



Opérations

- Adapter les associations items/correcteurs (non destructif)
- - modification de la grille d'évaluation
 - Recorrection des items "divergants"
 - Supprimer un les items d'un correcteur
 - RE corrige des items divergeants
- - RE Tranche (recorrige)
 - Recorrection des items "divergants"
 - Supprimer les items d'un correcteur
 - RE Tranche (selection score +fav)
 - Priorité correcteur
 - Supprimer les scores d'un correcteur

FIGURE 9 – Représentation schématisée du processus de correction et des diverses opérations de pilotage des corrections (RE signifie « Responsable d'évaluation ») proposées dans ExAMS dans le cas des Questions à réponses ouvertes longues (QROL)

Pour le travail de correction, les correcteurs bénéficient d'une interface originale dans laquelle la question est placée en parallèle de la grille de correction. En cas de besoin, le correcteur peut aussi faire appel à une réponse type encodée préalablement par le responsable de l'évaluation ou revoir une copie qu'il aurait marquée pour l'une ou l'autre raison (parce qu'elle est très bonne, par exemple).



FIGURE 10 – Interface de correction des Questions à réponses ouvertes longues (QROL) proposée dans ExAMS

Lorsque l'ensemble du travail est effectué, le responsable de l'évaluation clôture la phase de correction. La plateforme dispose ainsi des scores obtenus par les candidats pour chacun des indicateurs descriptifs du support. Il est alors possible de calculer une série de statistiques descriptives pour le test, les questions et les candidats (tout comme dans le cas des QCM). De plus, ces statistiques peuvent être globales (sur l'ensemble des questions de l'épreuve) ou calculées par catégories de contenus (par exemple les chapitres de la matière) ou par catégories de performances (par exemple les niveaux taxonomiques de Bloom) afin d'obtenir des tableaux de bord plus diagnostiques. Le responsable de l'évaluation pourrait être rapidement submergé par une grande quantité d'informations issues de ces tableaux de bord. C'est pourquoi le système propose de sélectionner et de visualiser les statistiques en fonction des besoins du responsable et, s'il le souhaite, de les imprimer.

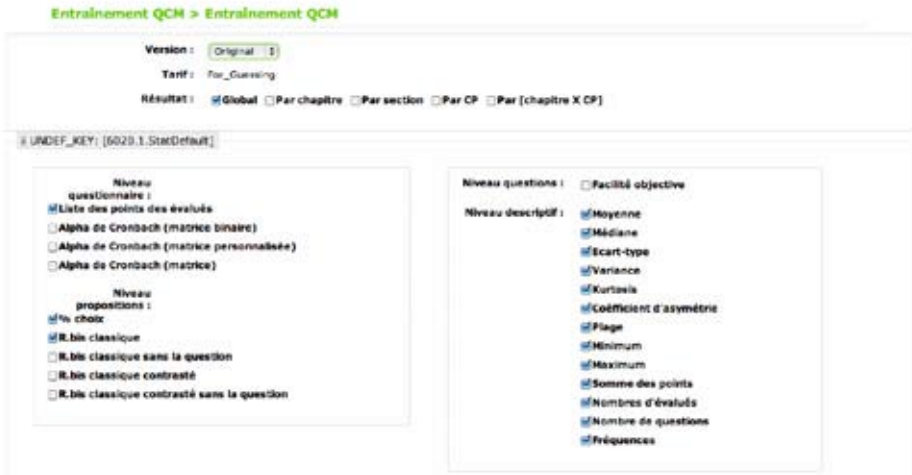


FIGURE 11 – Interface d'ExAMS permettant de sélectionner les statistiques qui figureront dans le tableau de bord d'une épreuve lors de la phase de correction

Le responsable de l'évaluation peut apporter certaines rectifications aux résultats après analyse des indices statistiques classiques tels que les histogrammes de fréquence des scores, les indices de cohérence interne... Dans le cas des QCM, les rectifications peuvent être effectuées sur trois éléments lors du « pilotage de qualité » : le questionnaire (par exemple, changement de tarif pour les calculs des scores), les questions (par exemple, suppression ou ajout d'une question) et la proposition (par exemple, modification de la réponse correcte).



FIGURE 12 – Interface d’ExAMS permettant de rectifier les résultats de l’épreuve lors du « pilotage de la qualité » de la phase de correction

Lorsqu’une version finale des scores de l’épreuve est entérinée après le « pilotage de la qualité », plus aucune modification ne peut être effectuée et le responsable de l’évaluation est alors en mesure de fournir les rétroactions aux candidats. Celles-ci peuvent être très diagnostiques si le travail d’analyse effectué lors de la première phase de l’élaboration a été réalisé consciencieusement. Des rétroactions précises peuvent alors être transmises aux candidats sur leur performance, chapitre par chapitre ou selon les catégories de performances visées.

Utilité d’ExAMS pour les candidats

Dès que le responsable de l’évaluation met à la disposition des candidats des informations relatives à l’épreuve, ces derniers reçoivent un courriel les informant de leur code d’accès à ExAMS. Ils peuvent alors se connecter et, en fonction des dates de disponibilité des informations, l’interface leur montre le contenu accessible en surbrillance et le reste en grisé. En sélectionnant une épreuve, le candidat a accès à l’ensemble de l’information disponible : documents associés, consignes, etc. Si le calendrier de mise à disposition le permet, il peut s’entraîner au moyen des épreuves de préparation ou encore passer l’épreuve certificative si le responsable de l’évaluation a prévu de la proposer en ligne.



FIGURE 13 – Interface d'ExAMS permettant de rectifier les résultats de l'épreuve lors du « pilotage de la qualité » de la phase de correction

Dans le cas d'une passation en ligne, l'interface de passation du test permet au candidat de naviguer d'une question à l'autre. Pour les QCM (figure 14 ci-dessous), un indicateur de l'état des questions apparaît sur l'interface. Le candidat peut sauter une question pour y revenir par la suite tant qu'il n'a pas validé sa réponse.

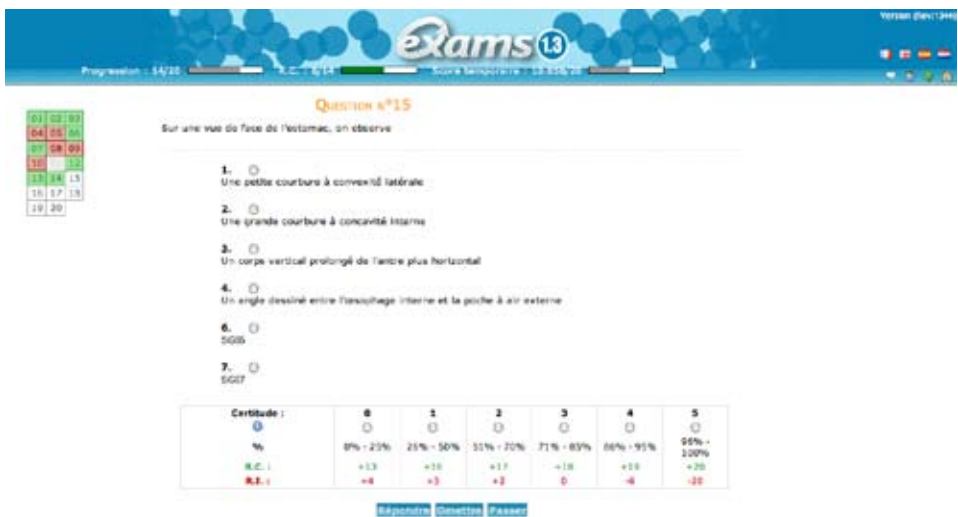


FIGURE 14 – Interface d'ExAMS pour les QCM

En fonction du paramétrage défini par le responsable de l'évaluation, l'étudiant peut se connecter à la plateforme pour accéder à ses rétroactions qu'elles soient issues d'une épreuve passée en ligne ou d'une épreuve sous forme papier-crayon.

CONCLUSIONS ET PERSPECTIVES

La plateforme ExAMS fournit au responsable de l'évaluation une aide précieuse. Elle lui offre une méthode structurée pour construire ses outils d'évaluation, elle lui permet de gérer ses banques de questions, de communiquer avec les candidats, de créer ses épreuves, d'en proposer des versions papier-crayon ou en ligne et de fournir rapidement des rétroactions diagnostiques sur les apprentissages réalisés par les candidats. Elle se base sur une méthodologie rigoureuse et structurée pour construire et gérer l'évaluation des apprentissages et offre des indices statistiques nécessaires aux prises de décisions docimologiques portant sur la régulation des épreuves, et ce, qu'elles soient basées sur des QCM ou des QROL. Pour ce faire, elle autorise le travail collaboratif supervisé par un responsable de l'évaluation et offre des possibilités d'intervention à distance pour des acteurs intervenant dans la création ou la validation des outils d'évaluation en construction. Le candidat quant à lui y trouve toute l'information nécessaire en vue de se préparer aux épreuves auxquelles il sera soumis. ExAMS gère des épreuves papier-crayon traditionnelles mais aussi des épreuves en ligne et, dans ce second cas, la plateforme offre une interface conviviale de passation. Enfin, que ce soit pour une épreuve en ligne ou papier-crayon, ExAMS permet au candidat d'accéder à des rétroactions diagnostiques.

ExAMS est une plateforme de code source libre (*open source*) téléchargeable gratuitement depuis le Web que chacun peut se réapproprier et adapter en fonction de ses besoins. Actuellement, une communauté de développeurs, de docimologues et d'utilisateurs se constitue progressivement autour d'ExAMS et toute personne intéressée est invitée à se joindre à cette communauté (Gilles et Ramelot 2010). Dans un avenir à moyen terme beaucoup de nouveautés seront apportées à la plateforme dans le cadre du projet QDFCA (*Quality Design For Competencies Assessment*) qui orientera l'outil vers la production d'épreuves dans une approche par compétences. Cela implique de renforcer les modalités de communication au sein de la plateforme entre les candidats et le responsable de l'évaluation en permettant plus d'interactions entre eux. Les performances des candidats devraient aussi pouvoir être observées sous plusieurs angles et pas seulement d'un point de vue purement cognitif. Dès lors, de nouvelles fonctionnalités multimédias seront offertes. Elles permettront de renforcer l'authenticité des situations d'évaluation et d'évaluer

de façon conviviale et rigoureuse des performances vidéoscopées en utilisant des grilles d'évaluation construites à l'aide de la plateforme.

Dans une approche par compétences, les standards de performance jouent un rôle important et l'implémentation récente dans ExAMS des supports descriptifs qui forcent à la description des critères, des aspects et des indicateurs de performances constitue une avancée dans cette logique des standards qui tend à décrire le plus précisément possible et en termes qualitatifs ce qui est attendu des candidats. En ce sens, les standards viennent renforcer la pratique des objectifs en décrivant les attentes pour chaque niveau à atteindre dans le processus de formation. ExAMS devra également renforcer l'intégration du recours aux jugements professionnels et permettre le recours à des jugements successifs qui devront ensuite être intégrés à un jugement global cohérent. Cela devrait amener de nouvelles possibilités de visualisation des profils de progression des candidats et permettre d'offrir à ces derniers de nouvelles possibilités de rétroactions diagnostiques. Il en découlera également des distinctions plus claires entre l'évaluation du processus et l'évaluation du produit. Beaucoup de savoirs, de savoir-faire et de savoir-être s'acquièrent et se construisent progressivement et ExAMS sera en mesure de montrer ces progressions autant que les productions finales. L'idée est d'offrir au sein de la plateforme la possibilité de combiner le jugement terminal au suivi de la progression dans le contexte d'une approche par compétences. De gros efforts devront également être effectués pour renforcer les liens entre les aspects enseignement et formation non pris en charge par la plateforme ExAMS, qui est un système de gestion des évaluations (*Assessment Management System [AMS]*), et d'autres types de plateformes plus orientées vers la gestion des apprentissages telles que les systèmes de gestion de l'apprentissage (*Learning Management System, [LMS]*). Ce type d'alliance entre un AMS et un LMS nous semble une voie d'avenir, car elle permettrait en effet d'envisager des liens directs entre des diagnostics dans une approche par compétences, gérés par l'AMS, et les mesures correctives liées aux situations d'apprentissage, gérées par le LMS. Il en résulterait également à travers ces liens entre diagnostics et remédiation un potentiel nouveau en termes de gestion informatisée des évaluations formatives dans une approche par compétences.

On le voit, c'est un ensemble de défis docimologiques et technologiques qui sont actuellement relevés par l'équipe du projet QDFCA et qui amènera à terme la plateforme ExAMS à franchir un nouveau palier dans son développement. Les bases méthodologiques du cycle de construction et de gestion de la qualité des outils pour l'évaluation sur lesquelles reposent la version actuelle ainsi que le paradigme de l'approche qualité contribueront au succès de ce projet qui vise à introduire l'approche par compétences dans ExAMS, et ce, en plaçant au centre de nos préoccupations les besoins des évaluateurs et des évalués.

Références

- American Educational Research Association, American Psychological Association, et National Council on Measurement in Education (dir.) (1999). *Standards for Educational and Psychological Testing*. Washington, DC, Etats-Unis, American Psychological Association.
- Blais, J.-G., M. Laurier, J.-M. Van der Maren, C. Gervais, M. Lévesque et G. Pelletier, (1997). *L'évaluation des apprentissages à l'Université de Montréal et dans ses écoles affiliées*. Montréal, Canada: Université de Montréal, Faculté des sciences de l'éducation, Groupe de recherche interdisciplinaire en pédagogie universitaire (GRIPU).
- Castaigne, J.-L., J.-L. Gilles et C. Hansen (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^e Congrès de l'Association Internationale de Pédagogie Universitaire (AIPU), Les stratégies de réussite dans l'enseignement supérieur*. Dakar, Sénégal, Université Cheikh Anta Diop.
- Chuan-Ju, L. (2008). « Comparisons between classical test theory and item response theory in automated assembly of parallel test forms ». *The Journal of Technology, Learning, and Assessment*, 6(8), p. 1-42.
- Cor, K., C. Alves et M. Gierl, M. (2009). « Three applications of automated test assembly within a user-friendly modelling environment ». *E-journal of Practical Assessment, Research and Evaluation*, 14(14).
- Crahay, V., E. Hausoul, R. Mainferme et J.-L. Gilles, (2006). « Vers une qualité documentaire accrue des évaluations des acquis des apprenants à l'aide d'un cycle de construction et de gestion des tests standardisés ». *23^e Congrès de l'Association Internationale de Pédagogie Universitaire (AIPU) Innovation, formation et recherche en pédagogie universitaire*. Monastir, Tunisie, Faculté de Pharmacie.
- Crahay, V. et J.-L. Gilles (2007). « Vers un accompagnement documentaire instrumenté à l'aide de la plateforme e-C&QCST pour les enseignants de l'Université de Liège qui pratiquent le *testing* standardisé ». *Actes du 24^e congrès de l'Association Internationale de Pédagogie Universitaire (AIPU). Vers un changement de culture en enseignement supérieur – Regards sur l'innovation, la collaboration et la valorisation*. Montréal, Canada, Université de Montréal.
- De Souza, E. et M. Fleming (2003). « A comparison of in-class and online quizzes on student exam performance ». *Journal of Computing in Higher Education* 14(2), p. 121-134.
- Downing, S. M. et T. M. Haladyna (1997). « Test item development: Validity evidence from quality assurance procedures ». *Applied Measurement in Education*, 10, p. 61-82.

- Downing, S. M. et T. M. Haladyna (2006). *Handbook of Test Development*. Mahwah, États-Unis: Lawrence Erlbaum Associates.
- Draaijer, S. et R. Hartog (2007). « Guidelines for the design of digital closed questions for assessment and learning in higher education ». *E-Journal of Instructional Science and Technology*, 10(1).
- Gilles, J.-L. et D. Leclercq (1995). « Procédures d'évaluation adaptées à des grands groupes d'étudiants universitaires – Enjeux et solutions pratiquées à la FAPSE-ULG ». *Symposium International sur la Rénovation Didactique en Biologie*. Tunis, Tunisie: Université de Tunis.
- 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, Université de Liège, Liège, Belgique.
- Gilles, J.-L. et V. Lovinfosse (2004). « Utilisation du cycle SMART de gestion qualité des évaluations standardisées dans le contexte d'une Haute École: Regard critique en termes de validité, fidélité, sensibilité des mesures, diagnosticité, praticabilité, équité, communicabilité et authenticité ». Dans S. Arzola, R. Vizcarra, A. Cornejo, G. Undurraga, (dir.), *World Education for Educational Research (WAER) – XIV Congreso Mundial de Ciencias de la Educación: Educadores para una Nueva Cultura – Summary of papers abstracts*. Santiago, Chili: Pontificia Universidad Católica de Chile, Facultad de Educación.
- Gilles, J.-L., S.-A. Piette, P. Detroz et M. Pirson (2004). « Le projet de plate-forme électronique de construction et de gestion qualité de tests standardisés e-C&QCST », Dans S. Arzola, R. Vizcarra, A. Cornejo, et G. Undurraga (dir.), *World education for educational research (WAER) – XIV Congreso Mundial de Ciencias de la Educación: Educadores para una Nueva Cultura – Summary of papers abstracts*. Santiago, Chili: Pontificia Universidad Católica de Chile, Facultad de Educación.
- Gilles, J.-L., S.-A. Piette, P. Detroz, S. Tinnirello, M. Pirson, M. Dabo et H. Lê (2005). « The electronic construction and quality control in standardized testing platform project e-C&QCST. » Dans A. Demetriou, F. Dochy, et collab. (dir.), *European Association For Research on Learning and Instruction (EARLI) – 11th Biennial Conference of EARLI: Book of Abstracts*. Nicosie, Chypre, University of Cyprus, Department of Educational Sciences.
- Gilles, J.-L., P. Detroz, V. Crahay, J.-P. Humblet, R. Mainferme et G. Radermaecker (2008a). « La plateforme Open-source ExAMS: une approche qualité pour le testing standardisé ». Dans F. Chafiqi et collab. (dir.), *Actes du 15^e congrès de l'Association Mondiale des Sciences de l'Éducation (AMSE) Mondialisation et éducation: vers une société de la connaissance* (Vol. I, p. 146-147). Marrakech, Maroc, Université Cadi Ayyad.
- Gilles, J.-L., P. Detroz, V. Crahay, J.-P. Humblet, Y. Lescalier, M. Pirson, S. Tinnirello, A. Mohr, H. Lê, R. Mainferme et G. Radermaecker (2008b). « Gérer et contrôler la qualité des évaluations standardisées à l'aide de la plate-forme open source ExAMS ». Dans F. Chafiqi, et collab. (dir.), *Actes du 15^e congrès de l'Association Mondiale des*

- Sciences de l'Éducation (AMSE) Mondialisation et éducation: Vers une société de la connaissance.* Marrakech, Maroc, Université Cadi Ayyad.
- Gilles, J.-L. et P. Detroz (2008). « L'enquête internationale PraDES au sein de l'Université de Liège: Premières leçons et impact sur le soutien aux enseignants ». Dans F. Chafiq et collab. (Dir.), *Actes du 15^e congrès de l'Association Mondiale des Sciences de l'Éducation (AMSE) Mondialisation et éducation: Vers une société de la connaissance.* Marrakech, Maroc, Université Cadi Ayyad.
- Gilles, J.-L. et P. Ramelot (2010). « Comment favoriser le partage d'expérience entre experts du terrain et chercheurs du monde académique pour faire progresser les connaissances en évaluation? ». Dans D. Castillo Moncada et collab. (dir.), *Actes du 16^e congrès de l'Association Mondiale des Sciences de l'Éducation (AMSE): Colloque évaluation de performances complexes.* Monterrey, Mexique, Escuela de Ciencias de la Educación.
- Gilles, J.-L. (2010). *Qualité spectrale des tests standardisés universitaires.* Sarrebruck, Allemagne: Éditions universitaires européennes.
- Jönsson, A., N. Mattheos, G. Svingby et R. Attström (2007a). Dynamic assessment and the « interactive examination ». *Educational Technology and Society*, 10, p. 17-27.
- Jönsson, A. et G. Svingby (2007b). « The use of scoring rubrics: Reliability, validity and educational consequences ». *Educational Research Review*, 2, p. 130-144.
- Kauffman, C., P. Dupont et A. Philippart (1995). *Projet pilote européen pour l'évaluation de la qualité dans l'enseignement supérieur.* Bruxelles, Belgique: Ministère de l'Éducation, de la Recherche et de la Formation.
- Kirkland, M. C. (1971). « The effects of tests on students and schools ». *Review of Educational Research*, 41, p. 303-350.
- Kulik, C.-L. et J. A. Kulik (1987). Mastery testing and student learning: A meta-analysis. *Journal of Educational Technology Systems* 15(3), p. 325-345.
- Leclercq, D. et J.-L. Gilles (1993). « Hypermedia: Teaching through assessment ». Dans D. Leclercq et J. Bruno (dir.), *Item Banking: Interactive Testing and Self Assessment (NATO ASI)*. (Vol. 112, p. 31-47). Berlin, Allemagne: Springer Verlag.
- Leclercq, D. et J.-L. Gilles (2001). « Techniques de mesure dans l'autoévaluation (dix techniques d'auto-estimation de la qualité de ses réponses) ». Dans G. Figari, et M. Achouche (dir.), *L'activité évaluative réinterrogée – Regards scolaires et socioprofessionnels* (p. 134-142). Bruxelles, Belgique: De Boeck.
- Nightingale, P. et M. O'Neil (1994). *Achieving quality in learning in higher education.* Londres, Grande-Bretagne: Kogan Page.
- Nitko, A. J. et S. M. Brookhart (2007). *Educational assessment of students* (5^e éd.). Upper Saddle River, États-Unis: Merrill-Prentice Hall.
- Phelps, R. P. (2005). *Defending standardized testing.* Mahwah, États-Unis: Lawrence Erlbaum.
- Ramsden, P. (1991). « A performance indicator of teaching quality in higher education: The course experience questionnaire ». *Studies in Higher Education*, 16, p. 129-150.

- Rasch, G. (1966). « An item analysis which takes individual differences into account ». *British Journal of Mathematical and Statistical Psychology*, 19(1), p. 49-57.
- Riopel, M., G. Raïche, M. Pilote et P. Potvin (2009). « La production automatisée de tâches d'évaluation ». Dans J.-G. Blais (dir.), *Évaluation des apprentissages et technologies de l'information et de la communication* (p. 257-274). Québec, Canada : Les Presses de l'Université Laval.
- Segers, M., F. Dochy et E. Cascallar (2003). « The era of assessment engineering: Changing perspectives on teaching and learning and the role of new modes of assessment ». Dans M. Segers, F. Dochy et E. Cascallar (dir.), *Optimising new modes of assessment: In search of qualities and standards* (p. 1-12). Dordrecht, Pays-Bas : Kluwer Academic.
- Stocking, M. L., R. Smith et L. Swanson (2000). *An investigation of approaches to computerizing the GRE subject tests*. Research Report 00-04. Princeton, États-Unis : Educational Testing Service.
- Shufford, A., A. Albert et N. E. Massengil (1966). « Admissible probability measurement procedures ». *Psychometrika*, 31, p. 125-145.
- Thompson, N.A. (2008). « A proposed framework of test administration methods ». *Journal of Applied Testing Technology*, 9(5), p. 1-17.
- Zink, K.J., et A. Schmidt (1995). « Measuring universities against the European quality award criteria ». *Total Quality Management*, 6(5-6), p. 547-561.