

Le Bulletin de l'ADMEE-Europe

JANVIER 2010



SOMMAIRE

1. EDITORIAL	2
2. SITE INTERNET DE L'ADMEE-EUROPE : DE NOUVELLES PRATIQUES	4
3. ARTICLE THEMATIQUE : QUELQUES REFLEXIONS AUTOUR DES EPREUVES D'EVALUATION DEVELOPPEES DANS LE CADRE DE L'APPROCHE PAR COMPETENCES	5
4. RECENSIONS D'OUVRAGES	20
5. ANNONCE DE COLLOQUES ET SEMINAIRES	21
6. CONSEIL D'ADMINISTRATION DE L'ADMEE-EUROPE	22

1. EDITORIAL

Meilleurs vœux à vous toutes et tous !

Ce Bulletin, sortant au mois de janvier 2010 à l'occasion du colloque de Braga, est d'abord l'occasion de vous adresser à toutes et tous, au nom du Conseil d'administration de l'ADMEE-Europe, nos meilleurs vœux pour la nouvelle année. Que celle-ci soit des plus enrichissantes et l'occasion d'échanges et de collaborations scientifiques, amicales et festives !

Articles thématiques dans le Bulletin : des débats actuels et passionnants !

Ce présent *Bulletin* est avant tout thématique, poursuivant le débat sur l'évaluation des compétences initié dans le précédent *Bulletin* par l'article de *Vincent Carette* sur l'approche par compétences contrastée à la pédagogie par objectifs. La contribution de *Christophe Dierendonck* et *Annick Fagnant* dans ce *Bulletin* soulève la question des outils d'évaluation des compétences par les situations complexes et par les épreuves standardisées. Les auteurs soulignent les logiques différentes des outils et démarches méthodologiques associées. Leur objectif est ambitieux : concilier l'évaluation par situations complexes et les modèles de réponse à l'item, dans le but de penser de nouvelles pistes en matière d'évaluation formative et de feedback diagnostique. Le débat est ouvert : il est passionnant !

Mon souhait, en tant que présidente, est que le *Bulletin* poursuive la publication d'articles de haute qualité et puisse ainsi contribuer à nourrir les réflexions scientifiques et les pratiques d'évaluation en éducation et formation. L'année 2010 a pour objectif, entre autres, de concevoir le *Bulletin* pour que celui-ci ait les moyens de proposer des articles destinés tant aux chercheurs qu'aux praticiens de notre Association, dans l'idée d'une complémentarité avec la revue *Mesure et évaluation en éducation*. Je me réjouis d'avoir l'occasion de vous en dire davantage une fois que le projet sera plus élaboré.

Classement de la revue *Mesure et évaluation en éducation* au rang A

L'année 2009 a été l'occasion, par l'intermédiaire d'*Anne Jorro*, d'une procédure de reconnaissance de la revue *Mesure et évaluation en éducation* au niveau de la France. Ainsi, après évaluation, **la revue a été classée au rang A**. Il s'agit d'une excellente nouvelle qui reconnaît la position internationale de la revue, la qualité de sa procédure d'expertise et de ses publications scientifiques. L'objectif est de poursuivre la valorisation de la revue au plan européen. Nos remerciements vont à *Anne Jorro* pour le suivi de ce dossier d'importance.

Une première année de fonctionnement du « nouveau » Conseil d'administration

Cette première année du Conseil d'administration 2009-2010 a été marquée par plusieurs autres dossiers importants. Au plan scientifique :

- L'organisation par nos collègues portugais Palmira Alves et André Machado, du 22ème colloque de l'ADMEE-Europe à Braga sur le thème de l'évaluation et du curriculum.
- La proposition par les délégations d'activités nationales très riches. Un compte-rendu des activités réalisées en 2009 sera présenté dans le prochain Bulletin, ainsi que les propositions pour 2010.
- La reprise de la rédaction européenne de la revue *Mesure et évaluation en éducation* par *Claude Houssemand* et la poursuite d'une collaboration de haute qualité avec l'ADMEE-Canada. Et bien sûr la procédure de reconnaissance de la revue mentionnée plus haut.

Au plan administratif, les dossiers ont concerné notamment :

- La réalisation du nouveau site Internet de l'Association et des nouveaux outils de gestion associés. La section II de ce *Bulletin* présente plus en détail ce dossier pour clarifier les modifications de pratiques engendrées.
- L'ajustement des pratiques administratives du secrétariat de l'ADMEE-Europe en lien avec les outils techniques mis en place.
- La professionnalisation du secrétariat depuis janvier 2009 qui est désormais rétribué par l'Association.

C'est ici l'occasion d'annoncer l'arrivée, en été 2009, d'une nouvelle collaboratrice au secrétariat de l'ADMEE-Europe, *Nathalie Nazzari* (info@admee-europe.org). Je lui souhaite la bienvenue et la remercie déjà pour la qualité de nos collaborations. Je profite aussi de remercier chaleureusement *Elisabeth Egger* qui a assuré le secrétariat courant du premier semestre 2009 et plus généralement l'IRDP, dirigé par *Matthis Behrens*, qui offre depuis plusieurs années ses compétences et l'infrastructure technique aux tâches administratives de l'Association. C'est extrêmement précieux.

L'Assemblée générale du vendredi 15 janvier 2010 à 17h30 communiquera plus en détail des informations sur ces dossiers et d'autres. Je vous rappelle qu'en tant que membres vous êtes cordialement invités à participer à cette Assemblée générale.

Cotisation 2010 et tarifs des différents types d'adhésion

Une nouvelle année annonce bien sûr la cotisation annuelle. L'Assemblée générale de 2009 à Louvain-la-Neuve, en raison notamment du secrétariat qui s'est professionnalisé, a voté une augmentation des cotisations dès 2010.

Les nouveaux tarifs sont les suivants :

Membre individuel :	65 € (+ 15 €)
Membre étudiant :	50 € (+ 10 €)
Membre institutionnel :	130 € (+ 20 €).

Ces tarifs comprennent :

- les numéros annuels de la revue *Mesure et évaluation en éducation*
- le *Bulletin* publié plusieurs fois par année
- un *tarif préférentiel* pour participer aux manifestations de l'Association
- l'accès au tout nouvel *Espace membre* du site Internet
- et, plus généralement en tant que membres de notre réseau international, vous serez régulièrement informés de l'actualité de la recherche en évaluation et de l'évolution des pratiques en ce domaine.

Sans prétention aucune, nous pouvons constater un développement constant de l'ADMEE-Europe, au travers des colloques internationaux, des activités nationales, des publications sous formes d'actes et d'ouvrages, ainsi qu'au regard des analyses et comptes-rendus de pratiques, un axe praxéologique que l'Association a toujours valorisé et qu'elle souhaite continuer de promouvoir. Ce développement n'a été possible que grâce au soutien de ses membres, autrement dit grâce à vous. Nous vous en remercions sincèrement et espérons pouvoir compter sur votre fidélité.

Remerciements à Nathalie Droyer, rédactrice sortante du Bulletin

Enfin, je tiens à terminer cet éditorial en remerciant chaleureusement la rédactrice sortante du *Bulletin*, *Nathalie Droyer*. Elle a assumé avec compétence pendant plusieurs années la double fonction de déléguée nationale pour la France et de rédactrice du *Bulletin*. Au nom de tous, je te souhaite, Nathalie, une très belle suite !

Lucie Mottier Lopez, présidente

2. SITE INTERNET DE L'ADMEE-EUROPE : DE NOUVELLES PRATIQUES

Comme j'ai eu l'occasion de le dire, un travail acharné (et le mot n'est pas excessif !) a été réalisé pour la mise en œuvre du nouveau site Internet de l'ADMEE-Europe, grâce notamment au soutien financier et technique de la société *Cinformatique* en Suisse.

Pour plus d'efficacité, j'ai accepté d'assumer la fonction de webmaster, le temps que le site soit totalement opérationnel tant au plan technique qu'au niveau des contenus. En effet, le changement opéré est d'importance, conduisant à une transformation des pratiques pour le secrétariat de l'ADMEE-Europe et de sa gestion des membres, pour les membres eux-mêmes et pour les délégations nationales. L'objectif, ici, est d'expliquer les principaux changements.

Paiement des cotisations en ligne : www.admee.org

- Le paiement des cotisations annuelles doit *obligatoirement* passer par une procédure en ligne qui remplace l'ancien formulaire papier.
- Cette procédure concerne tant une *nouvelle* adhésion qu'un *renouvellement* d'une adhésion.
- Pour ce faire, il faut cliquer sur la rubrique « Membre, devenir membre », puis compléter les renseignements personnels.
- Il s'agit ensuite de procéder au paiement en ligne qui est totalement sécurisé (Paypal).
- Une option de versement bancaire est encore proposée mais uniquement pour des cas particuliers. Dans ce cas, il faut aussi compléter les renseignements personnels dans les rubriques en ligne.
- Notre *objectif* est que progressivement tous les versements puissent se réaliser par le moyen de Paypal.
- Cette *démarche* doit se faire *chaque année*, y compris quand il s'agit d'un renouvellement d'une adhésion.
- Elle remplace avantageusement, pour les membres et pour le secrétariat, l'envoi postal de l'ancien formulaire d'adhésion.
- Par cette nouvelle pratique, la base de données des membres (en ligne) est mise à jour automatiquement avec les renseignements que vous remplissez.
- Votre cotisation est effective à la *réception du versement*. Il ne suffit donc pas de remplir les rubriques en ligne pour être membre.

Une gestion partagée du site Internet

La conception professionnelle du site Internet, sa base de données intégrée et les modules techniques mis en place permettent une gestion partagée des contenus et de leur actualisation.

- Chaque *délégation nationale* a un accès aux pages Internet qui la concernent et elle est *responsable de la mise en ligne des informations* sur les *activités nationales* proposées et de les maintenir à jour.
- Le secrétariat a un accès à toutes les pages qui concernent les informations générales de l'Association.
- Enfin, le webmaster gère la structure du site, sa cohérence et sa transformation au fur et à mesure des besoins. Il s'occupe également de l'espace membre.

Un « espace membre » : une nouvelle offre aux membres de l'ADMEE-Europe

- Dès 2010, un espace offrira des informations spécialement destinées aux membres de l'ADMEE-Europe. Il s'agit d'une *nouvelle offre* qui vise à promouvoir et à alimenter les échanges au sein de l'Association.
- Dans cet espace, nous envisageons notamment des *articles thématiques*, la possibilité pour les membres de soumettre une contribution qui peut aussi être un partage de pratiques, l'accès aux différents documents propres au fonctionnement associatif. Cet espace sera géré par le webmaster.
- J'espère que cet espace sera opérationnalisé courant du premier trimestre 2010. Chaque membre recevra alors un identifiant et mot de passe pour y accéder, en cliquant sur « Membre, connexion ».

Je ne peux ainsi que vous inciter à consulter régulièrement notre site qui est maintenant *l'outil principal de communication* de l'ADMEE-Europe !

Lucie Mottier Lopez, présidente



3. ARTICLE THEMATIQUE : QUELQUES REFLEXIONS AUTOUR DES EPREUVES D'EVALUATION DEVELOPPEES DANS LE CADRE DE L'APPROCHE PAR COMPETENCES

1. Introduction

Dans le précédent bulletin de l'ADMEE (2009/2), Vincent Carette nous a offert une intéressante confrontation : pédagogie par objectifs vs approche par compétences. Après avoir clairement mis en lumière l'influence behavioriste de l'une et la parenté constructiviste de l'autre¹, il conclut son texte en parlant de continuité entre les deux approches au niveau des buts poursuivis (définir un curriculum structuré qui permet la programmation et l'évaluation des apprentissages) mais de rupture au niveau de ce qui est attendu des élèves : « [...] dans les pédagogies par objectifs, il est attendu que les élèves soient capables de connaître, de comprendre, d'appliquer, d'analyser, de synthétiser (pour reprendre les niveaux des activités intellectuelles de Bloom) ; dans une pédagogie par les compétences, il est attendu que les élèves soient capables de mobiliser, de transférer, de mettre en œuvre un ensemble organisé de savoirs, de savoir-faire qui leur permettent d'accomplir une tâche » (p. 24). Cette rupture nécessaire qui amène à penser tous les apprentissages dans une perspective de mobilisation plaide selon lui pour le développement de nouveaux outils d'évaluation adaptés à cette approche par les compétences.

Au départ d'une analyse critique de deux nouveaux outils d'évaluation développés dans le cadre de l'approche par compétences - le modèle d'évaluation de Rey, Carette, Defrance et Kahn (2003) et les épreuves d'évaluation par situations complexes de De Ketele et Gérard (2005) - le présent article interroge la possibilité d'intégrer certains apports de ces outils au sein des évaluations standardisées classiques des acquis des élèves.

¹ Jonnaert (2002) suggère lui aussi que la logique de l'approche par compétences s'ancre dans la perspective socioconstructiviste. Si on peut déceler une certaine parenté entre les objectifs de l'approche par compétences et certains fondements de la perspective socioconstructiviste, Crahay et Forget (2006) se refusent pourtant à y voir un déterminisme propre aux sciences de l'éducation, qui résulterait d'un cheminement autonome de la pensée pédagogique. Selon ces auteurs, il s'agit plutôt d'examiner les enjeux sociétaux qui peuvent expliquer l'exportation d'une notion, issue du monde de l'entreprise, sur le terrain de l'éducation (voir aussi Crahay, 2006).

2. Le modèle d'évaluation des compétences de Rey et al. (2003)

2.1. Présentation du modèle

Rey et al. (2003) définissent la compétence comme « le fait de savoir accomplir efficacement une tâche, c'est-à-dire une action ayant un but » (p. 33). L'« authentique compétence » serait « la capacité à répondre à des situations complexes et inédites par une combinaison nouvelle de procédures connues ; et non pas seulement à répondre par une procédure stéréotypée à un signal préétabli » (p. 26). A partir de cette définition, les auteurs distinguent trois degrés de compétences qu'ils définissent comme suit :

- Les compétences de premier degré (« compétences élémentaires », « procédures », « procédures élémentaires » ou « procédures automatisées ») : « Savoir exécuter une opération (ou une suite prédéterminée d'opérations) en réponse à un signal (qui peut être en classe une question, une consigne, ou une situation connue et identifiable sans difficulté, ni ambiguïté) » (p. 26).
- Les compétences de deuxième degré (« compétences élémentaires avec cadrage ») : « Posséder toute une gamme de ces compétences élémentaires et savoir, dans une situation inédite, choisir celle qui convient » (p. 26). Dans ce type de situation, une interprétation de la situation (ou un « cadrage » selon la terminologie des auteurs) est nécessaire. L'élève doit pouvoir déterminer quelle est la procédure qui convient, parmi l'ensemble de celles qu'il a apprises. Autrement dit, il doit mobiliser (et non simplement appliquer) la procédure adéquate.
- Les compétences de troisième degré (« compétences complexes ») : « Savoir choisir et combiner plusieurs compétences élémentaires pour traiter une situation nouvelle et complexe » (p. 26). Dans ce type de situation, une interprétation de la situation et une organisation de la démarche de résolution sont nécessaires. En effet, ces tâches évaluent « la capacité des élèves à choisir et à combiner, parmi les procédures qu'ils connaissent, plusieurs d'entre elles afin de résoudre adéquatement un problème nouveau pour eux » (p. 44).

Sur la base de cette distinction, les auteurs développent un modèle d'évaluation des compétences en trois phases, synthétisé dans le tableau suivant.

Tableau 1 : Les trois degrés de compétence et le modèle d'évaluation en trois phases définis par Rey et al. (2003)

Degrés de compétence	Modèle d'évaluation des compétences	
Compétence de 3 ^e degré ou compétence complexe	Phase 1	« On propose aux élèves une tâche complexe exigeant le choix et la combinaison d'un nombre significatif de procédures que les élèves doivent posséder. Il est préférable que cette tâche soit pluridisciplinaire et fonctionnelle » (p. 92).
Compétence de 2 ^e degré ou compétence élémentaire avec cadrage	Phase 2	« On propose la même tâche complexe, mais découpée en tâches élémentaires, présentées dans l'ordre où elles doivent être accomplies pour réaliser la tâche globale. L'élève doit choisir, parmi les procédures qu'il connaît, celle qui convient à chacune des tâches élémentaires » (p. 92).
Compétence de 1 ^{er} degré ou compétence élémentaire ou procédure	Phase 3	« On demande aux élèves d'accomplir les procédures exigées dans les phases précédentes, mais sous une forme décontextualisée [...] » (p. 92).

Dans leur ouvrage, Rey et al. (2003) concrétisent leur modèle d'évaluation des compétences en décrivant trois épreuves testées auprès d'élèves de 3^e année primaire et trois épreuves proposées en 6^e année primaire. L'expérimentation de ces épreuves auprès de deux échantillons d'élèves de 8 ans et 12 ans permet aux auteurs de dresser quatre constats à partir des pourcentages de réussite observés aux différentes phases du modèle : (a) la majorité des élèves ne parviennent pas à résoudre les tâches complexes de la phase 1, (b) il existe une hiérarchie de réussite entre les phases (la réussite augmente fortement entre la phase 1, la phase 2 et la phase 3), (c) la cohérence du modèle semble éprouvée (la réussite d'un élève à la phase 1 est étroitement liée à sa réussite en phase 2 et en phase 3) et (d) la maîtrise des procédures automatisées (phase 3) est une condition nécessaire mais pas suffisante pour réussir les tâches élémentaires avec cadrage (phase 2) et *a fortiori* les tâches complexes (phase 1).

2.2. Discussion du modèle

La discussion est organisée en trois points : le premier s'interroge sur le découpage « surprise » des tâches présentées lors de la phase 1 du modèle, le deuxième questionne son caractère strictement hiérarchique et le troisième porte sur sa faisabilité dans des épreuves d'évaluation externes nationales ou internationales, tout en questionnant la fiabilité des informations récoltées.

2.2.1. Le découpage « surprise » de la phase 1

Quand on analyse les six épreuves utilisées par les auteurs pour expérimenter leur modèle, on s'aperçoit lors de la phase 1, que les tâches globales initiales sont déjà découpées en trois sous-tâches. Ceci étonne au vu de la définition que les auteurs donnent de la première phase de leur modèle (cf. tableau 1). Pour justifier ce découpage inattendu, la raison suivante est avancée :

Pratiquement, il est souvent nécessaire que la tâche globale proposée dans la phase 1 de l'épreuve soit pré-découpée, dès cette phase, en deux ou trois sous-tâches. D'une certaine manière, un tel découpage paraît contrevenir au caractère global et complexe que doit avoir cette tâche. Mais il est pratiquement indispensable, faute de quoi la tâche de départ peut donner lieu à des stratégies de résolution très diverses, parmi lesquelles certaines pourraient bien contourner l'usage de toute procédure apprise à l'école. L'indication explicite de deux ou trois sous-tâches auxquelles doit donner lieu la tâche globale permet de prévenir de telles dérives (Rey et al., 2003, p. 49).

A la lecture de cette justification, deux questions s'imposent : (1) le fait de proposer aux élèves une tâche globale complexe sans un pré-découpage en deux ou trois sous-tâches serait-il voué à l'échec? Et (2) la raison pratique avancée par les auteurs ne cache-t-elle pas l'existence d'un saut qualitatif entre les tâches annoncées de la phase 1 et les tâches pré-découpées effectivement proposées aux élèves? En effet, si l'on considère que les sous-tâches proposées évaluent des compétences de troisième degré, quels types de compétences la tâche globale non découpée évalue-t-elle finalement ? En d'autres termes, ne faudrait-il pas envisager, sur le plan théorique *a minima*, l'existence d'un niveau de compétence supplémentaire avec des compétences de quatrième degré qui seraient quant à elles nécessairement multi-tâches et/ou pluridisciplinaires?

2.2.2. Les phases du modèle d'évaluation sont-elles strictement hiérarchiques ?

Les données présentées par Rey et al. (2003, p. 104) fournissent assurément la preuve empirique qu'il existe une certaine hiérarchie entre les trois phases de leur modèle d'évaluation. En moyenne, les élèves réussissent mieux les items de phase 2 (tâches décomposées) que les items de phase 1 (tâches complexes) tandis que les items proposés en phase 3 (évaluation des procédures de base nécessaires aux autres phases) sont de loin les items les mieux réussis.

Rien ne permet cependant d'affirmer que la hiérarchie découverte est une hiérarchie stricte. Certains élèves pourraient échouer à certains items de la phase 3 (compétences élémentaires ou procédures) tout en ayant passé avec succès les phases 1 et 2. Comment cela pourrait-il s'expliquer ? Tout d'abord, les items proposés en phase 3 par les

auteurs dépassent (assez largement parfois) le cadre de l'application stricte des procédures mobilisées aux phases 1 et 2. Par exemple, si une multiplication intervient lors de la phase 1, la phase 3 évalue plus largement cette procédure en proposant des items impliquant différents niveaux de maîtrise (par exemple² : $38 \times 33 = 209 \times \dots$). Si on reconnaît volontiers l'intérêt de proposer une évaluation plus large des procédures (nous y reviendrons d'ailleurs par la suite), force est de constater que cela questionne la hiérarchie stricte du modèle. En effet, certains items de la phase 3 peuvent présenter un niveau de difficulté tel qu'ils soient moins bien réussis qu'un item de phase 2 ou même de phase 1.

La hiérarchie entre les phases du modèle peut également être questionnée dans un autre cas de figure. On peut par exemple imaginer qu'un élève réussisse un item présenté dans un certain contexte (par exemple : calculer le périmètre d'une zone de jeu pour y placer une clôture) parce que ce contexte lui permet de bien distinguer les notions d'aire et de périmètre (voire de développer une démarche informelle qui n'est pas basée sur les formules apprises) alors que dans une situation décontextualisée (par exemple : calculer le périmètre de la figure suivante), ce même élève échoue parce que, placé dans une telle situation formelle, il confond simplement les formules d'aire et de périmètre. Cela dit, même si un tel exemple casse la hiérarchie évoquée (l'élève aurait par exemple réussi la phase 2, tout en échouant à un item de phase 3), le modèle demeure intéressant d'un point de vue diagnostique.

Ces deux exemples illustrent assez bien selon nous la nuance qu'il convient de faire entre les notions de difficulté et de complexité. Il est en effet possible qu'une tâche complexe (au sens de nécessitant la mobilisation et l'intégration de plusieurs ressources/procédures) présente un niveau de difficulté relativement faible et qu'à l'inverse, une tâche élémentaire s'avère extrêmement difficile pour les élèves. Tâche complexe et tâche compliquée (pour reprendre la terminologie de Roegiers, 2007) ne sont en effet nullement synonymes. Nous y reviendrons par la suite, lorsque nous reprendrons la question d'une évaluation suffisamment large des ressources/procédures et de leur mobilisation/application dans des tâches élémentaires.

2.2.3. La faisabilité d'un tel modèle d'évaluation et la fiabilité des informations récoltées

Au-delà de la considération éthique qui suppose d'évaluer uniquement ce qui a fait l'objet d'un apprentissage formel, on peut avancer, en reprenant les limites que les auteurs eux-mêmes ont formulées, que le modèle d'évaluation proposé est difficilement envisageable dans le cadre d'une épreuve externe locale, nationale ou internationale, qui devrait se dérouler en un temps limité (généralement 2 à 3 heures). En effet, au vu de la complexité des tâches présentées en phase 1 et des tâches qui en découlent aux phases 2 et 3, il semble difficile de pouvoir proposer plus de deux ou trois tâches complexes et pluridisciplinaires par épreuve d'évaluation. Du point de vue de la mesure, cela pose question, surtout si l'on veut décider du degré de compétence des élèves dans une discipline donnée sur la seule base de l'épreuve proposée.

Sur les plans conceptuel et diagnostique, l'apport du modèle d'évaluation de Rey et al. (2003) est réel puisque, théoriquement, il permet non seulement de savoir si les élèves sont capables de résoudre des tâches complexes mais il autorise également l'examen des lacunes des élèves en ciblant précisément la phase où les élèves ont échoué. La démarche d'évaluation proposée souffre cependant d'une faiblesse liée à la fiabilité de la mesure puisque les interprétations ou les décisions (d'ordre pédagogique ou autres) qui pourraient en découler ne sont, en réalité, basées que sur un nombre restreint d'items. Or, quel que soit le modèle de mesure utilisé (théorie classique des tests ou théorie de réponse à l'item), plus on dispose d'items qui évaluent ce que l'on souhaite mesurer, plus la mesure que l'on prend devrait être précise. C'est le principe de la théorie de l'échantillonnage : dans la théorie classique des tests, plus un échantillon est grand, plus l'estimation des caractéristiques de l'univers de référence ou la population dont il est tiré (ici, l'ensemble des items que l'on pourrait produire pour évaluer une compétence) tend à être précise puisque la somme des erreurs aléatoires de mesure tend vers zéro lors-

² Exemple tiré d'une des épreuves d'évaluation développée par Rey et al. (2003, p. 84).

qu'un grand nombre d'items est utilisé (Laveault et Grégoire, 1997). Des études réalisées sur les modèles de réponse à l'item montrent par ailleurs l'influence du nombre d'items sur la précision des estimations de la compétence des sujets (Hulin, Lissak et Drasgow, 1982 ; Burton, 2004). Conscients de cette difficulté, inhérente selon eux aux nouveaux outils d'évaluation proposés, les auteurs font appel à « de nouvelles approches statistiques qui restent à ce jour à inventer » (Carette, 2009, p. 155).

3. Les épreuves d'évaluation par situations complexes de De Ketele et Gérard

3.1. Présentation des épreuves d'évaluation par situations complexes

Pour De Ketele et Gérard (2005), l'approche par compétences cherche à développer la possibilité par les apprenants de mobiliser un ensemble intégré de ressources pour résoudre une situation-problème appartenant à une famille de situations³. Selon eux, les épreuves élaborées selon l'approche par compétences consistent, par essence, « à présenter à l'élève une, voire deux situations complexes, demandant de la part de l'élève une production elle-même complexe, nécessitant un certain temps de résolution » (p. 8). Gérard (2008) affirme que l'élément le plus important dans la notion de compétence est l'aspect « intégration » : « la compétence se manifeste dans une situation complexe qui intègre un certain nombre d'éléments et qui nécessite de mobiliser – c'est-à-dire d'identifier et d'utiliser conjointement – un ensemble intégré de ressources » (p. 51).

Au niveau de la construction de situations d'évaluation, les auteurs défendent « avec force le principe que la pertinence est première par rapport à la validité et à la fiabilité. Ces dernières sont évidemment importantes à nos yeux, mais on ne peut s'en contenter [...] » (p.22). S'appuyant sur cet argument, De Ketele et Gérard (2005) critiquent la pertinence des épreuves d'évaluation telles que TIMSS ou PISA :

Ces épreuves proposent un ensemble d'items selon une structure très élaborée issue principalement d'une approche par les contenus ou par les objectifs. Elles sont fiables et valides, mais ne sont pas pertinentes par rapport à l'approche par compétences et donc par rapport aux finalités des systèmes éducatifs conçus dans cet esprit. Si elles permettent bien d'évaluer les ressources jugées nécessaires (savoir-reproduire et savoir-faire), elles ne permettent pas (ou peu) d'évaluer la faculté de mobiliser celles qui sont pertinentes pour résoudre des problèmes ou effectuer des tâches complexes (De Ketele & Gérard, 2005, p. 5).

Pour correspondre à leur définition de la compétence, De Ketele et Gérard (2005) estiment que les tâches demandées aux élèves peuvent être de deux ordres : soit on propose une tâche unique complexe (ex : produire un texte répondant à une situation de communication), soit on décompose une production en plusieurs tâches comprenant elles-mêmes plusieurs étapes. Mais, précisent les chercheurs, « en tous les cas, il ne sera pas possible de constituer un nombre d' « items », parce que ceux-ci reviendraient à décomposer la tâche complexe en sous-unités, ce qui ne permettrait plus d'évaluer la compétence qui consiste bien à gérer la complexité » (p. 8). Ceci constitue l'argument principal des auteurs pour justifier la déclaration suivante :

Il n'est donc pas question, dans des épreuves d'évaluation construites selon l'approche par les compétences, de valider les outils comme on le fait avec des épreuves classiques [...] en prenant comme unités d'analyse les « items » constituant l'échantillon issu de l'univers de référence. Dans une épreuve selon les compétences, ces 'items' n'existent pas : il n'y a qu'une production complexe, ne constituant certainement pas un échantillon statistiquement représentatif de toutes les productions complexes possibles relativement à la compétence. Les techniques de validation des épreuves d'évaluation selon l'approche par compétences sont donc à inventer, avec d'autres postulats que ceux de la théorie classique des scores (De Ketele & Gérard, pp. 8-9).

De Ketele et Gérard (2005) proposent trois types de validation des épreuves d'évaluation selon l'approche par compétences : une validation a priori par recours à des juges, une

³ A ce propos, Gérard (2008) signale que « tout le monde ne s'accorde pas sur la notion de 'familles de situations'. Certains chercheurs en contestent même la validité (Crahay, 2005, 2006 ; Rey, Carette, Defrance et Kahn, 2003) » (p. 61).

validation empirique interne et une validation empirique externe. La validation a priori par recours à des juges consiste, dans le cas de l'approche par compétences, à demander à des juges de vérifier que les situations concrètes (les « items » dans l'approche classique) qui sont proposées aux élèves font bien partie de la famille de situations (l'« univers de référence » dans l'approche classique) qui correspond à la compétence visée. La validation empirique interne vise à vérifier l'unité conceptuelle de l'épreuve en recourant aux techniques de l'analyse factorielle et aux coefficients d'homogénéité. Quant à la validation empirique externe, elle nécessite de recourir à des critères externes qui seront définis à partir d'études comparatives (ex : groupe expérimental - groupe contrôle) ou d'analyses corrélationnelles nécessitant des mesures répétées.

Dans un ouvrage à vocation pratique, Gérard (2008) propose une démarche en sept étapes pour élaborer une situation évaluant une compétence : (1) définir les objectifs de l'évaluation et la démarche à adopter, (2) déterminer avec précision, pour une discipline et un niveau, la compétence (ou le palier de compétence) que l'on va évaluer à partir de la situation, (3) élaborer une situation d'intégration qui correspond à la compétence visée, (4) s'assurer que les paramètres de la situation d'évaluation correspondent bien à ceux de la famille de situations prévues pour la compétence, (5) vérifier si la situation mobilise des ressources adéquates à la compétence et préalablement apprises, (6) recourir aux critères d'évaluation de la compétence et faire en sorte que l'épreuve permette de les évaluer au moins trois fois de manière indépendante (règle des 2/3) et déterminer les indicateurs qui conviennent et (7) s'assurer que la situation d'évaluation en a bien toutes les caractéristiques.

3.2. Discussion des épreuves d'évaluation par situations complexes

La discussion est organisée en deux points : le premier interroge la façon dont les auteurs interprètent la notion de « pertinence » des épreuves d'évaluation et le deuxième questionne le degré de fiabilité des informations qu'ils récoltent dans le cadre des évaluations par situations complexes.

3.2.1. Une interprétation discutable de la notion de « pertinence » des épreuves

De Ketele & Gérard (2005) définissent la pertinence comme étant « le caractère plus ou moins approprié de l'épreuve, selon qu'elle s'inscrit dans la ligne des objectifs visés » (p. 3). Pour être pertinentes dans une approche par les compétences, les épreuves d'évaluation devraient donc mesurer de réelles compétences. Selon eux, « la compétence se manifeste dans une situation complexe qui intègre un certain nombre d'éléments et qui nécessite de mobiliser – c'est-à-dire d'identifier et d'utiliser conjointement – un ensemble intégré de ressources » (Gérard, 2008, p. 51), Cette définition correspond assez bien aux compétences de troisième degré définies par Rey et al. (2003). Si l'on peut s'accorder globalement sur cette définition, nous aurions néanmoins envie de soutenir que des compétences peuvent aussi se manifester dans des situations moins complexes, pour autant qu'elles nécessitent, *a minima*, la mobilisation de ressources ou de procédures, et non leur simple application (ce qui correspond alors aux compétences de deuxième degré selon la définition de Rey et al., 2003). De plus, on voit mal pourquoi les concepts de « situation complexe » et d'« ensemble intégré de ressources » devraient nécessairement conduire à des situations « très » complexes et intégrant « un nombre important de ressources ». Plus concrètement, nous voyons difficilement pourquoi certaines questions de l'épreuve PISA (la plupart même) ne pourraient pas être considérées comme évaluant des compétences (de deuxième degré à tout le moins, mais parfois aussi de troisième degré). Dans la plupart des questions de l'épreuve PISA, les élèves sont en effet placés face à des situations (des problèmes à résoudre) qui nécessitent la mobilisation d'une ou plusieurs ressources ou procédures élémentaires et qui dépassent clairement leur simple application. Autrement dit, déclarer qu'un élève situé au niveau 5 de l'échelle évaluant la culture mathématique dans PISA fait preuve de compétences en mathématiques ne nous semble nullement « impertinent ».

3.2.2. A propos de la fiabilité des informations récoltées

On s'accordera à dire qu'il n'est pas intéressant de disposer d'épreuves pertinentes et valides si ces dernières ne sont pas, en même temps, fiables. Plus précisément, le reproche que nous adressons aux évaluations par situations complexes telles qu'elles sont définies par De Ketele et Gérard (2005) et Gérard (2008) repose sur l'idée qu'on peut évaluer une compétence à partir de deux ou trois situations complexes, de critères d'évaluation (les qualités minimales que doit respecter la production de l'élève) et de la « règle des 2/3 » (De Ketele, 1996). Cette règle, « validée empiriquement » selon Roegiers (2005, p. 7), postule d'offrir à l'élève au moins trois occasions indépendantes de démontrer sa compétence. Selon Gérard (2008), « on considèrera qu'un élève maîtrise un critère lorsqu'il réussit au moins 2 des 3 occasions qui lui sont offertes » (p. 80). Il semblerait donc suffisant que l'élève réussisse deux tâches complexes sur trois pour être déclaré compétent, ce qui interpelle en termes de fiabilité de la mesure car cela voudrait également dire qu'il suffit de 2 ou 3 items pour évaluer de manière fiable un comportement ou la maîtrise d'une ressource ou d'une procédure précise. On rejoint ici la critique faite à l'encontre du modèle d'évaluation de Rey et al. (2003) à partir de la théorie de l'échantillonnage.

4. Peut-on s'inspirer des nouveaux outils d'évaluation pour enrichir les évaluations standardisées classiques ?

La quête d'outils alternatifs d'évaluation adaptés à l'approche par compétences intéresse de plus en plus de chercheurs francophones (Carette, 2007a, b ; De Ketele et Gérard, 2005 ; Gérard, 2008 ; Rey, Carette, Defrance et Kahn, 2003 ; Scallon, 2004). Avec Scallon (2004), on dira que cette volonté de disposer de nouvelles pratiques d'évaluation n'est pas propre à la francophonie puisqu'elle est apparue⁴ également, et initialement, aux Etats-Unis au milieu des années 1990 sous l'influence du courant *Competency-Based Education* caractérisé notamment par le rejet des tests standardisés classiques (composés de questions à correction objective de type vrai-faux, questions à choix multiple et/ou questions à réponse brève) et par le développement de la notion de *performance assessment* (appréciation de la performance). Cette notion suppose d'évaluer les compétences des élèves dans le cadre de situations complexes et authentiques.

Ce sont donc ces notions de *performance assessment* et de compétence en situation qui ont conduit au développement d'outils de mesure originaux dont il a fallu questionner la validité. Face à ce défi, d'aucuns ont écarté implicitement ou explicitement les procédures classiques de validation afin d'éprouver leurs modèles, préférant inventer de nouvelles démarches d'expérimentation ou de validation. C'est le cas, selon nous, du modèle d'évaluation en trois phases développé par Rey et al. (2003) et des épreuves d'évaluation par situations complexes proposées par De Ketele et Gérard (2005) et Gérard (2008).

Or, la question au centre du débat revient à se demander s'il faut rejeter d'emblée toute possibilité de conciliation entre les procédures classiques de validation et l'évaluation de compétences ?

4.1. Deux logiques vraiment inconciliables ?

Du côté des épreuves d'évaluation par situations complexes, on cherche avant tout à être en adéquation avec une approche pédagogique particulière, en l'occurrence l'approche par compétences. Ces épreuves se caractérisent aussi par leur volonté d'être diagnostiques et s'inscrivent clairement dans une perspective éducatrice (identifier les lacunes de chaque élève pour pouvoir y remédier).

⁴ Selon Hébert, Valois et Frenette (2008), « malgré des rapprochements certains entre les approches anglo-saxonne (la situation de performance) et francophone (la situation de compétence), la conception francophone actuelle du concept de situation pour apprécier des compétences semble avoir progressé indépendamment du courant anglo-saxon, qui est de toute façon très peu cité dans les écrits francophones ».

Du côté de la théorie de réponse à l'item (TRI), la perspective est plutôt psychométrique. Il s'agit de situer, à l'aide d'un modèle probabiliste, des individus et des items sur une échelle commune censée traduire à la fois la compétence des sujets (le trait latent) et la difficulté des items, avec l'avantage que les caractéristiques des items sont considérées comme indépendantes des caractéristiques des sujets et inversement, si le modèle s'ajuste correctement aux données. Une des particularités des modèles de réponse à l'item est le postulat d'unidimensionnalité qui suppose que tous les items d'un test doivent mesurer un seul et unique trait latent. Pour Laveault et Grégoire (1997), ce critère n'est jamais parfaitement rencontré dans la pratique du fait des inévitables erreurs de mesure et de la complexité des traits mesurés. Une autre particularité des modèles de réponse à l'item est qu'ils postulent l'indépendance des items : une fois le trait latent mesuré, aucune relation ne doit exister entre les réponses d'un sujet aux différents items. Cette condition d'application risque donc d'entrer en conflit direct avec les modèles d'évaluation qui prônent un découpage en phases et qui suggèrent une certaine dépendance entre les items.

Vouloir concilier l'évaluation par situations complexes et les modèles de réponse à l'item revient donc à tenter le mariage du feu et de l'eau. Nous pensons pourtant que cette union forcée est possible à la condition de trouver un terrain d'entente entre les deux parties. De cet exercice pourraient naître de nouvelles pistes en matière d'évaluation formative et de feedback diagnostique.

C'est dans cet esprit que s'ancrent les réflexions qui suivent. Elles sont avant tout formulées pour être soumises au débat contradictoire et à d'éventuelles tentatives de validation empirique. Elles n'ont pas d'autre prétention.

4.2. A la recherche d'un terrain d'entente...

La conciliation entre les nouveaux outils d'évaluation des compétences et les épreuves classiques d'évaluation pourrait selon nous trouver un terrain d'entente moyennant deux conditions essentielles : s'accorder sur une définition moins ambitieuse des tâches complexes et s'assurer que les épreuves comportent suffisamment d'items⁵.

4.2.1. S'accorder sur une définition moins ambitieuse des tâches complexes

Dans les deux modèles d'évaluation discutés jusqu'ici, les tâches ou les situations complexes sont définies à partir d'un niveau de complexité ultime : sont complexes les tâches nécessitant la mobilisation et l'intégration d'un grand nombre de ressources internes ou externes ; les tâches pluridisciplinaires étant considérées comme les mieux à même de remplir ces critères. Rappelons toutefois que ces définitions sont théoriques et que la concrétisation des épreuves d'évaluation conduit parfois à proposer des sous-tâches qui ne sont plus vraiment pluridisciplinaires et sans doute un peu moins complexes qu'attendu (cf. le découpage « surprise » de la phase 1 dans le modèle de Rey et al., 2003).

Théoriquement, on peut toujours imaginer soumettre aux élèves un nombre suffisant de tâches ou de situations complexes pour s'assurer d'une mesure fiable des compétences visées. Le problème se pose davantage en termes pratiques puisqu'il n'est pas envisageable d'organiser des séances d'évaluation de plusieurs heures qui seraient nécessaires à la résolution d'un nombre significatif de tâches ou de situations complexes. On l'a déjà dit, en deux heures de temps, on peut au maximum soumettre deux ou trois tâches complexes aux élèves.

⁵ Précisons que ce dernier point est d'emblée jugé par certains comme étant incompatible avec l'approche par compétences. Pour De Ketele et Gérard (2005), les items n'existent pas puisque l'on a affaire à une production complexe à analyser. Pour Carrette (2009), la déclinaison d'une épreuve en items mesurant des objectifs isolés n'aurait aucun sens puisque « ce que l'on désire évaluer, c'est la capacité qu'a un élève, dans une situation donnée, à choisir et à combiner des procédures normalement apprises » et que, selon lui, « la caractéristique essentielle d'une épreuve d'évaluation de compétences est 'de mettre l'élève devant une feuille blanche' » (p. 154). Hormis les problèmes de correction et ceux inhérents à la durée de résolution des tâches complexes, on voit mal pourquoi évaluer l'élève devant une feuille blanche (c'est-à-dire sans guider sa démarche par de multiples sous-questions) serait en soi contradictoire avec la possibilité de construire une épreuve comprenant de nombreux items.

Nous suggérons dès lors d'adopter une définition plus raisonnable des tâches ou situations complexes : une tâche/une situation serait considérée comme complexe à partir du moment où elle nécessite l'identification, la mobilisation et l'intégration de plus d'une ressource et qu'elle nécessite dès lors une interprétation (ou un cadrage) de la situation et une organisation de la démarche de résolution⁶.

Signalons que les données empiriques récoltées par Rey et al. (2003) appuient en partie cette décision d'abandonner, dans le cas qui nous occupe, les tâches ou les situations complexes de très haut niveau puisqu'on constate que très peu d'élèves (de 3 à 5%) parviennent à résoudre, avec un seuil de réussite de 80%, les deux tâches complexes proposées au sein de leur épreuve d'évaluation. Au seuil de réussite de 50%, ils ne sont que 21 à 46% selon l'épreuve considérée, ce qui confirme que la majorité des élèves éprouvent de très grandes difficultés face aux tâches complexes de la phase 1. Compte tenu de ces chiffres, on peut donc s'interroger sur la pertinence de proposer des tâches aussi complexes puisque d'une part, seuls très peu d'élèves s'y sentent à l'aise⁷ et que, d'autre part, le pouvoir diagnostique de la phase 1 reste limité puisque peu fiable du point de vue de la mesure.

Les épreuves construites à partir d'une définition raisonnable de la complexité seront sans doute considérées par Rey et al. (2003) et par De Ketele et Gérard (2005) comme moins « pertinentes » que les épreuves complexes pluridisciplinaires ou d'intégration qu'ils défendent en faisant référence à ce qu'ils semblent considérer comme le but ultime de l'approche par les compétences, mais elles permettront tout de même d'évaluer la compétence des élèves à résoudre des tâches complexes, et ce, même si le degré de complexité envisagé est plus faible.

4.2.2. S'assurer que les épreuves comportent suffisamment d'items

A partir de cette définition moins ambitieuse des tâches complexes, il devient possible de construire des évaluations constituées d'un plus grand nombre d'items évaluant des compétences complexes et de s'assurer ainsi d'une mesure plus fiable. Proposer des épreuves standardisées comportant un nombre suffisant d'items semble en effet être un impératif dicté par les théories de la mesure. Selon nous, il ne s'agit pas seulement d'une contrainte psychométrique mais également d'un impératif au niveau diagnostique.

La distinction opérée par Roegiers (2007) entre tâche compliquée et tâche complexe permet d'illustrer notre propos. Selon cet auteur, une tâche compliquée pour l'élève est une tâche « qui mobilise des acquis nouveaux, peu ou pas connus par lui, insuffisamment maîtrisés, ou qui lui sont peu familiers » (p. 16). La complexité mettrait en jeu quelque chose d'assez différent dans la mesure où elle ne dépendrait pas tellement du type de savoir, de savoir-faire ou de savoir être à mobiliser, mais surtout de leur quantité : « la difficulté vient non pas de chaque opération à exécuter, mais de l'articulation de ces opérations entre elles » (p. 121). Autrement dit, on pourrait imaginer élaborer des tâches complexes qui ne mobilisent en fait que des savoirs et des savoir-faire élémentaires, ce qui pourrait d'ailleurs se justifier pour éviter de mêler toutes les difficultés en même temps (procédures compliquées à articuler de façon complexe). Dès lors, si on propose, suivant le modèle de Rey et al. (2003), une épreuve en trois phases où les élèves sont amenés à résoudre une tâche complexe (phase 1), puis cette même tâche décomposée en tâches élémentaires (phase 2), puis enfin des questions visant l'application isolée des ressources/procédures mobilisées dans les étapes précédentes (phase 3), on risque un appauvrissement des épreuves en termes de contenus évalués, tant en termes de savoirs que de savoir-faire. Cela, Rey et al. (2003) semblent l'avoir compris puisque dans leurs exemples d'épreuves, ils contournent ce problème potentiel en proposant en phase 3 une évaluation des ressources/procédures qui dépasse l'application isolée de la ressource ou de la procédure impliquée dans les phases précédentes.

⁶ Soulignons que cette définition n'entre pas en contradiction avec la définition des compétences de troisième degré de Rey et al., 2003 (savoir choisir et combiner correctement plusieurs procédures de base pour traiter une situation nouvelle et complexe).

⁷ Ce n'est évidemment pas une raison pour éviter de proposer ce genre de tâches aux élèves en situation d'apprentissage voire lors d'une évaluation ponctuelle qui viendrait en complément d'autres outils d'évaluation (épreuves standardisées, portfolio, ...).

Concrètement, pour éviter le double écueil qui consisterait soit à proposer des situations complexes trop compliquées, soit à négliger l'évaluation de savoirs et de savoir-faire compliqués (ce qui constituerait selon nous une erreur d'un point de vue diagnostique), nous défendons l'idée que les évaluations standardisées devraient également contenir, en plus de quelques tâches complexes évaluant des compétences complexes (les compétences de troisième degré selon Rey et al. 2003 et les seules compétences réelles selon De Ketele & Gérard, 2005), un nombre suffisant de tâches élémentaires permettant d'évaluer la mobilisation de certaines ressources/procédures ainsi que leur application directe (compétences de deuxième degré et compétences élémentaires ou « procédures » selon Rey et al., 2003).

Si certaines de ces tâches élémentaires peuvent être liées à la tâche complexe (selon le phasage proposé par Rey et al., 2003), il nous semble important que d'autres en soient totalement indépendantes, et ce pour deux raisons principales : d'une part, cela permet d'évaluer la mobilisation et l'application d'une plus grande diversité de ressources/procédures que celles directement impliquées dans la tâche complexe initiale⁸ et, d'autre part, on donne ainsi plus de chances aux élèves de démontrer leur maîtrise des compétences de deuxième degré en leur proposant des contextes diversifiés. On peut en effet craindre qu'un élève qui échoue lors de la phase 1 du modèle de Rey et al. (2003) échoue aussi lors de la phase 2 parce qu'il est déconcerté par la situation proposée ou, tout simplement, parce qu'il se sent découragé. Ce même élève résoudrait peut être correctement d'autres tâches de complexité similaire, évaluant les mêmes ressources/procédures, mais présentées dans un autre contexte.

4.3. Quelques pistes de réflexion ... à partir d'une épreuve fictive

Il ne nous appartient pas de proposer ici un modèle d'épreuve d'évaluation qui colle à la fois aux exigences de l'approche par compétences et des modèles de réponse à l'item. Tout au plus, nous allons nous permettre d'évoquer certaines questions ouvertes et d'amorcer la discussion de certaines pistes qui pourraient déboucher sur une forme d'évaluation originale. Notre réflexion en la matière en est à ses balbutiements et notre propos se limitera à pointer les pierres d'achoppement d'une telle démarche et à suggérer l'une ou l'autre piste de travail qu'il conviendrait de valider ou d'invalider à la lumière d'études empiriques ou de considérations théoriques en lien avec la théorie de réponse à l'item. Nous articulons ces questionnements et ces pistes de travail autour d'un exemple fictif d'épreuve d'évaluation.

4.3.1. Une épreuve fictive comme point de départ

Pour donner une base concrète à notre réflexion, nous allons tenter d'imaginer une épreuve qui viserait à enrichir les épreuves standardisées classiques par certains apports des nouveaux outils d'évaluation. Plus précisément, nous voulons élaborer une épreuve fictive qui combinerait les éléments suivants :

- un nombre suffisant d'items pour assurer la fiabilité des informations récoltées ;
- une évaluation de la compétence des élèves à résoudre des tâches complexes (selon notre acception du terme) et des tâches élémentaires présentées dans différents contextes ou hors contexte ;
- une évaluation diagnostique de la compétence des élèves à résoudre des tâches complexes en utilisant un phasage similaire à l'approche de Rey et al. (2003) ;
- l'utilisation de modèles de réponse à l'item.

Supposons que l'objectif soit de réaliser une évaluation diagnostique des compétences des élèves en mathématiques. La première étape consiste évidemment à préciser quels domaines mathématiques, quelles compétences et quelles ressources/procédures sont visés par l'épreuve sachant qu'il est impossible de tout évaluer au travers d'une seule épreuve d'évaluation. Il est donc nécessaire d'opérer un choix au sein du référentiel de

⁸ Les tâches élémentaires pourraient ainsi non seulement impliquer une plus grande diversité de savoirs et savoir-faire mais aussi se permettre de faire appel à différents niveaux de maîtrise (tâches plus ou moins compliquées selon la terminologie de Roegiers, 2007).

compétences officiel. Dans la présente épreuve fictive, il s'agirait d'évaluer les compétences des élèves en résolution de problèmes dans les deux domaines suivants : nombres et opérations d'une part et géométrie d'autre part. L'épreuve poursuivrait également un objectif diagnostique : pouvoir identifier les raisons qui ont conduit certains élèves à échouer face aux tâches complexes présentées.

Concrètement, l'épreuve serait constituée de deux parties, chacune composée d'un type de tâches particulier. Dans la première partie, il y aurait par exemple 10 tâches complexes se déclinant potentiellement en trois phases d'évaluation. Dans la seconde partie, on proposerait par exemple une quarantaine d'items correspondant à des tâches élémentaires évaluant l'application de diverses ressources/procédures et leur mobilisation dans des contextes diversifiés. Cette déclinaison des tâches complexes en trois phases d'évaluation s'inspire clairement du modèle de Rey et al. (2003), tout en différant sur trois points essentiels. Une première différence est que la phase 1 devrait donner lieu à une simple évaluation de type réponse correcte / réponse incorrecte⁹. La deuxième différence est qu'en phase 2, lorsque diverses sous-questions sont posées pour guider la démarche de résolution, on ne tient compte à nouveau que de la réponse finale à la tâche de façon à ne disposer que d'un seul item¹⁰. Enfin, la dernière différence est qu'en phase 3, seuls les items évaluant la maîtrise isolée des ressources/procédures impliquées lors des phases 1 et 2 sont proposés (il n'y aurait donc pas, dans cette première partie de l'épreuve fictive, d'évaluation plus large des ressources/procédures comme c'est le cas dans les épreuves proposées par Rey et al., 2003).

4.3.2. Quel est l'intérêt de proposer des tâches complexes qui se déclinent en différentes phases d'évaluation ?

L'intérêt du modèle de Rey et al. (2003) réside, nous l'avons dit, dans le potentiel diagnostique du phasage qu'il propose. En fonction des réponses des élèves aux phases 1, 2 et 3, il est possible d'identifier précisément à quelle étape de résolution de la tâche les élèves éprouvent des difficultés. Si l'élève réussit une tâche complexe en phase 1, c'est qu'il a pu définir une démarche correcte de résolution et qu'il a été capable de mobiliser et d'intégrer les ressources/procédures nécessaires. S'il échoue en phase 1 mais qu'il réussit en phase 2, c'est qu'il éprouve sans doute quelques difficultés en termes de démarche de résolution puisque les ressources semblent quant à elles maîtrisées. Si l'élève échoue en phase 1 et en phase 2, on lui soumet uniquement des tâches élémentaires qui interviennent dans la tâche complexe. S'il les réussit, on peut supposer que son problème se situe plutôt sur les plans de la mobilisation et de l'intégration adéquates des ressources/procédures. S'il échoue à ces items, c'est que ses difficultés se situent, avant tout, au niveau de la maîtrise isolée des ressources/procédures.

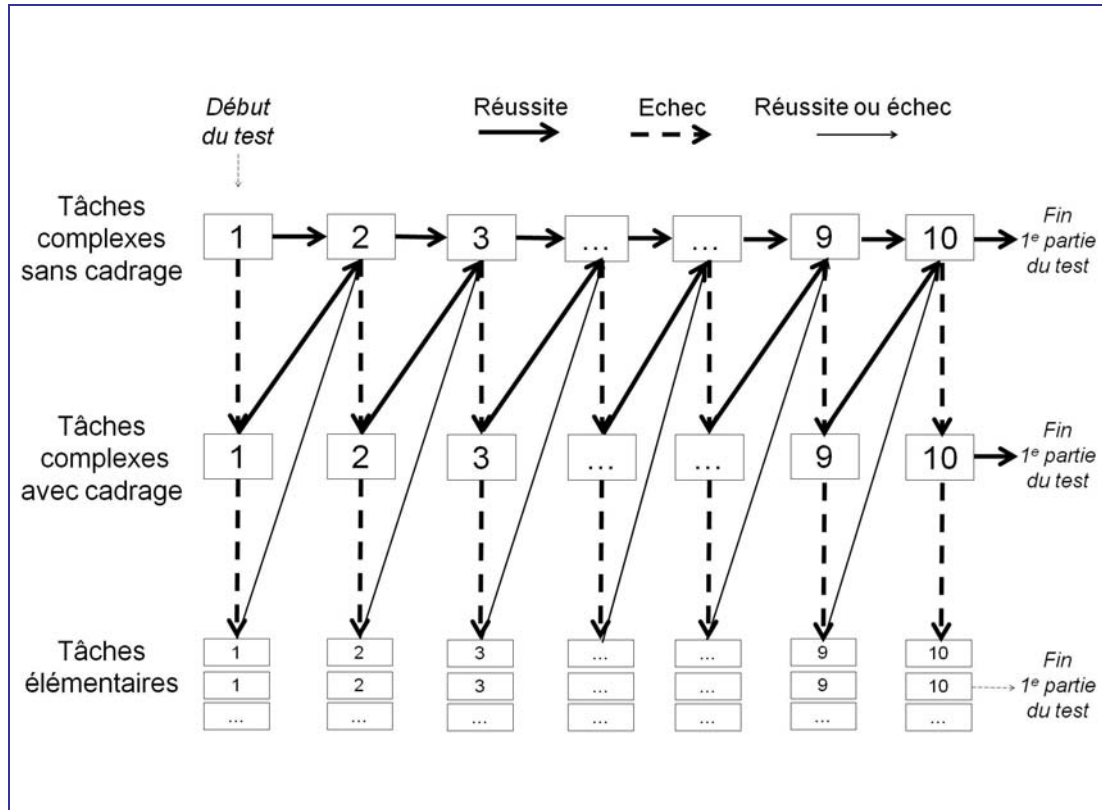
Dans notre exemple fictif, on proposerait 10 tâches complexes fidèles à cette logique. Grâce à une procédure de testing assisté par ordinateur, on pourrait imaginer ne présenter à l'élève que les items qui sont nécessaires pour déterminer son niveau de compétence et identifier où se situent ses lacunes. Pratiquement, pour cette première partie de l'épreuve fictive, la démarche serait la suivante (figure 1). Dans un premier temps, l'ordinateur soumet à l'élève la première tâche complexe, et ce, sans aucun cadrage (phase 1). En cas de réussite, l'ordinateur propose la deuxième tâche complexe. En cas d'échec, l'ordinateur présente à nouveau la première tâche complexe mais, cette fois, accompagnée d'un cadrage, en l'occurrence une démarche possible de résolution qu'il « suffit » de suivre pour obtenir la réponse correcte (phase 2). Si, au terme de la

⁹ Rappelons que nous situons ici nos propos dans le cadre d'une épreuve standardisée fictive, qui devrait être corrigée de façon centralisée ou de façon identique par tous les enseignants, ce qui impose à la fois une certaine rigueur dans les critères de correction, mais aussi une certaine simplicité. Par ailleurs, l'idée est que la perte d'informations liée à ce type de correction serait en partie compensée par les informations obtenues grâce aux résultats observés lors des phases ultérieures de la démarche. Evidemment, la situation est toute autre si l'enseignant utilise une épreuve diagnostique dans sa classe pour faire un bilan interne sur le niveau d'avancement de ses élèves. Dans cette optique, il peut recourir à des critères d'évaluation plus complexes, comme ceux proposés notamment dans les outils d'évaluation proposés par Rey et al. (2003) ou construits sur la base de leur modèle (voir le site [Hwww.enseignement.be](http://www.enseignement.be) pour quelques exemples d'outils utilisables dans les classes à des fins diagnostiques).

¹⁰ Se pose en phase 2 la question de savoir s'il est plus pertinent de proposer une nouvelle série de tâches à résoudre (dont les solutions successives conduisent à la résolution de la tâche complexe initialement proposée en phase 1) ou de lister simplement les différentes étapes d'une démarche à accomplir pour mener à bien la résolution de la tâche complexe ? La question de savoir si ces deux manières de procéder relèvent ou non d'un même niveau de compétences reste totalement ouverte.

phase 2, l'élève trouve la bonne réponse, l'ordinateur présente la deuxième tâche complexe. Si l'élève échoue en phase 2, l'ordinateur soumet les items évaluant la maîtrise isolée des ressources/procédures qui sont nécessaires à la résolution de la tâche complexe de départ (phase 3).

Figure 1 : Schématisation de la procédure de testing adaptatif par ordinateur utilisée dans la première partie de l'épreuve fictive



Cette procédure adaptative est assez éloignée de ce qui se fait classiquement avec les modèles de réponse à l'item (voir Laveault et Grégoire, 1997, p. 306). Ici, à un moment ou à un autre, l'ordinateur propose à tous les élèves chacune des 10 tâches complexes sans cadrage. Les autres items sont présentés uniquement en cas d'échec aux tâches de phase 1 ou de phase 2.

4.3.3. Comment concilier un modèle d'évaluation en phases et le postulat d'indépendance des items dans la théorie de réponse à l'item ?

Cette question est sans aucun doute la plus épineuse dans le cas qui nous occupe. Elle pose en tous les cas un défi majeur aux partisans des modèles de réponse à l'item. En effet, la TRI postule que les items doivent être strictement indépendants les uns des autres. Or, le simple fait de proposer un phasage ou un testing adaptatif au niveau de l'évaluation conduit, par définition, à proposer des items qui, dans une certaine mesure, sont liés au sein d'une même tâche complexe.

Face à ce problème de taille, recourir à un modèle à crédit partiel pourrait peut-être s'avérer une piste intéressante ? On recoderait ainsi chaque tâche complexe et son éventuel phasage en un seul item à crédit partiel (par exemple, quatre points si la tâche est réussie en phase 1, trois points si elle est réussie en phase 2, deux points si tous les items de la phase 3 sont réussis, un point pour une réussite partielle en phase 3 et aucun point en cas d'échec à tous les items)¹¹.

¹¹ Nous sommes bien conscients que cette proposition de codage mérite d'être améliorée dans la mesure où l'on perd de l'information dès qu'il y a un ou plusieurs échecs face aux tâches élémentaires puisque le recodage en crédit partiel ne permet plus de cibler face à quel(s) item(s) précisément l'élève a échoué.

4.3.4. A propos du postulat d'unidimensionnalité

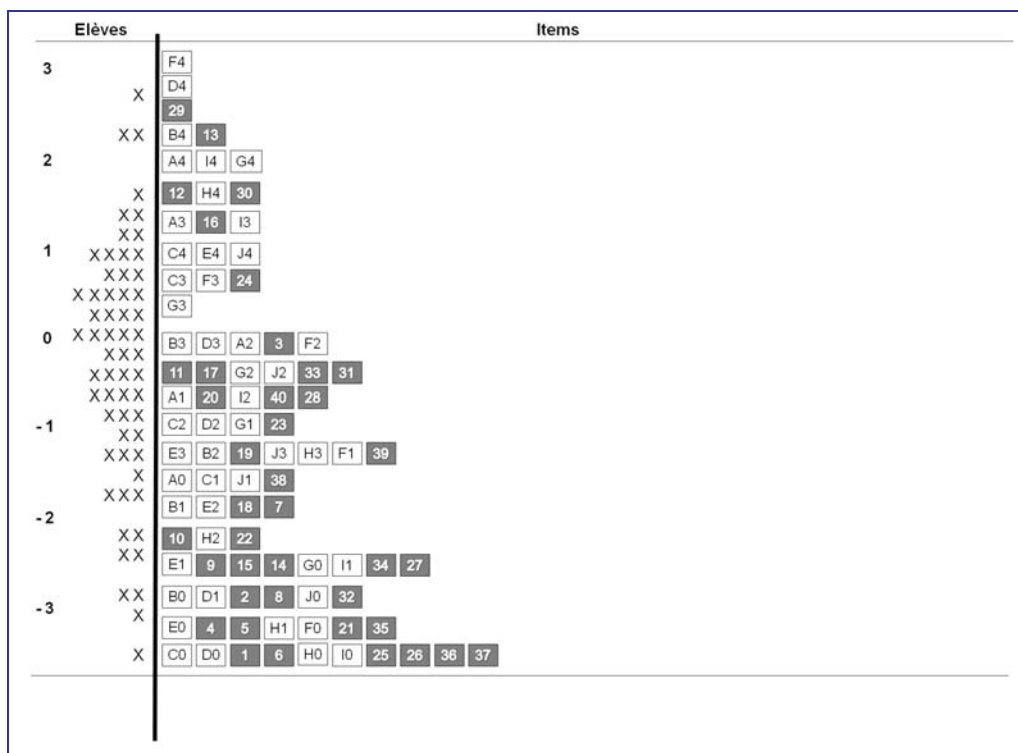
L'idée précédente de recoder en crédit partiel des tâches complexes pour contourner le problème de la dépendance des items nous amène à une autre question fondamentale : celle de savoir ce que l'on mesure exactement. Sur la base des différentes phases d'évaluation, on peut en effet se demander si on mesure « la même chose » à des niveaux de maîtrise différents ou si le fait de distinguer trois phases d'évaluation ne revient pas à mesurer trois traits latents distincts. Si on opte pour la seconde alternative, faut-il se diriger vers trois modélisations unidimensionnelles (dont on pourrait étudier les liens a posteriori) ou plutôt vers une modélisation multidimensionnelle ? Cette question de recherche est fondamentale à nos yeux mais nous n'avons actuellement aucune réponse à apporter.

4.3.5. Opter pour une présentation des résultats qui soit plus diagnostique

Le dernier élément que nous voulons soumettre au débat porte sur la valeur ajoutée des modèles de réponses à l'item par rapport à la théorie classique des scores en termes d'informations diagnostiques.

Rappelons-le une fois encore, la TRI permet de situer, au moyen d'un modèle probabiliste, des individus et des items sur une échelle commune censée traduire à la fois la compétence des sujets et les caractéristiques des items, tout en postulant que les caractéristiques des items sont indépendantes des caractéristiques des sujets et inversement. Au départ de notre épreuve fictive, il est donc possible d'obtenir une représentation graphique similaire à la figure 2 rendant compte à la fois de la distribution des élèves en termes de compétence et de la distribution des items en termes de difficulté. Les 10 items à crédit partiel sont représentés par une lettre (de A à O) et par un chiffre (crédit partiel : 4, 3, 2, 1 ou 0) tandis que les 40 tâches élémentaires sont représentées par un chiffre (de 1 à 40).

Figure 2 : Distribution fictive de la compétence des sujets et de la difficulté des items dans le cadre de notre épreuve de mathématiques (présentation classique)

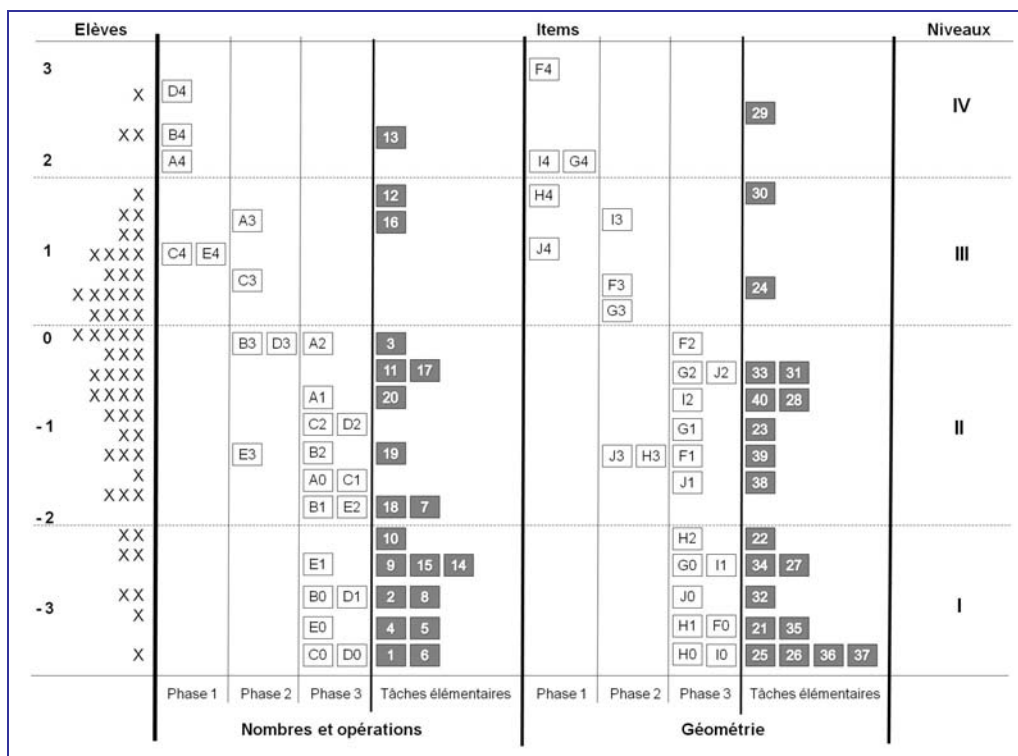


Ce graphique pourrait être adapté à des fins diagnostiques (figure 3). Au départ, le modèle de réponse à l'item renseigne uniquement la compétence de chaque sujet et l'indice de difficulté de chaque item sur une même échelle (axe vertical à la gauche du graphique). Il serait à notre avis intéressant de réaliser *a posteriori* un travail de description

des items. Ainsi, nous nous sommes limités à isoler les items selon leur appartenance aux domaines évalués (nombres et opérations ou géométrie) et aux phases d'évaluation (phase 1, phase 2, phase 3 ou tâches élémentaires). D'autres choix auraient pu être opérés. Par exemple, nous aurions pu distinguer les 40 items de la seconde partie de l'épreuve selon le type de comportement attendu (mobilisation ou simple application d'une ressource/procédure).

En outre, nous avons distingué à la droite du graphique quatre niveaux de compétence dont la logique est similaire aux niveaux de compétence que l'on trouve dans les échelles PISA. Ces niveaux de compétence sont également élaborés *a posteriori* soit par une analyse qualitative des items, soit par une analyse factorielle, soit en faisant référence à des seuils de performance, soit via une autre méthode à définir. Tout l'art consiste cependant à définir des niveaux de compétence qui présentent une cohérence et une logique propres et qui correspondent effectivement aux items censés les décrire.

Figure 3 : Distribution fictive de la compétence des sujets et de la difficulté des items dans le cadre de notre épreuve de mathématiques (présentation diagnostique)



Les retombées diagnostiques d'un tel graphique nous semblent évidentes. On dispose non seulement des informations normatives classiques qui permettent d'identifier les élèves les plus compétents et ceux qui éprouvent de grandes difficultés mais il est possible - et c'est là que se situe la valeur ajoutée des modèles de réponse à l'item - d'associer le degré de compétence de chaque élève face à des tâches mathématiques de complexité variable.

Dans notre exemple, on identifie assez bien quatre groupes d'élèves qui présentent des forces et des faiblesses différentes et qui pourraient faire l'objet d'une approche pédagogique différenciée. Reste alors à savoir si les épreuves standardisées peuvent s'arrêter à de tels constats ou si l'on devrait également fournir des pistes didactiques aux enseignants ?

5. Pour terminer ... sans conclure

Jette-t-on un pavé dans la mare en interrogeant la possibilité de concilier les évaluations standardisées classiques (basées de plus en plus sur les modèles de réponses à l'item) et les nouveaux outils d'évaluation des compétences ?

D'aucuns diront que oui et qu'à force de vouloir simplifier les situations dans lesquelles les élèves peuvent démontrer leurs compétences, on retombe dans les « travers » des évaluations construites sur la base d'une pédagogie par objectifs. D'autres penseront peut-être que des compromis raisonnables doivent être recherchés afin de ne pas ériger en norme la complexité inédite (pour paraphraser Crahay, 2006). Mais au-delà des clivages idéologiques, c'est sans doute à une définition encore plus précise de la notion même de compétence qu'il convient de faire appel. Pour Gérard (2008), « la compétence se manifeste dans une situation complexe qui intègre un certain nombre d'éléments et qui nécessite de mobiliser – c'est-à-dire d'identifier et d'utiliser conjointement – un ensemble intégré de ressources » (p. 51). Pour Carette (2009), dans une épreuve d'évaluation des compétences, « ce que l'on désire évaluer, c'est la capacité qu'a un élève, dans une situation donnée, à choisir et à combiner des procédures normalement apprises » (p. 154). Est-il réellement cohérent (ou pertinent) de préciser les définitions en identifiant notamment le nombre de ressources ou de procédures élémentaires à intégrer pour qu'une situation fasse réellement appel à la notion de compétence ? Ne gagnerait-on pas plus en pouvoir diagnostique (et en pertinence) en définissant différents degrés de compétence (comme ont tenté de l'initier Rey et al., 2003) et en cherchant à faire porter les évaluations sur ces différents degrés ? Les débats sur ces questions sont vives à l'heure actuelle, comme en témoignent les écrits de nombreux spécialistes sur ces questions (voir les ouvrages récents sur le sujet et notamment celui de Mottier Lopez & Crahay, 2009). La présente contribution avait pour objectif de susciter quelques réflexions, questionnements ou débats... Nous espérons que tel est le cas.

Références bibliographiques

- Burton, R. (2004). Influence des distributions du trait latent et de la difficulté des items sur les estimations du modèle de Birnbaum : une étude du type Monte-Carlo. *Mesure et évaluation en éducation*, 27, 3, 41-62.
- Carette, V. (2007a). L'évaluation au service de la gestion des paradoxes liés à la notion de compétence. *Mesure et évaluation*, 30(2), 49-71.
- Carette, V. (2007b). Les implications de la notion de compétence sur l'évaluation. *Éducation et formation*, e-286, 51-61.
- Carette, V. (2008). Pédagogie par objectifs et approche par compétences : rupture ou continuité ?, *Bulletin de l'ADMEE* 2009/2, 11-25.
- Carette, V. (2009). Et si on évaluait des compétences en classe ? A la recherche du cadrage instruit. In L. Mottier Lopez & M. Crahay (Eds.), *Evaluations en tension. Entre la régulation des apprentissages et le pilotage des systèmes* (149-163), Bruxelles : De Boeck.
- Crahay, M. (2006). Dangers, incertitudes et incomplétude de la logique de compétence. *Revue française de pédagogie*, 154, 97-110.
- Crahay, M. & Forget, A. (2006). Changements curriculaires : quelle est l'influence de l'économique et du politique ? In F. Audigier, M. Crahay & J. Dolz (Eds.), *Curriculum, enseignement et pilotage* (63-84). Bruxelles : De Boeck & Larcier.
- De Ketele, J.-M. & Gérard, F.-M. (2005). La validation des épreuves d'évaluation selon l'approche par les compétences, *Mesure et évaluation en éducation*, 28(3), 1-26.
- Gérard, F.-M., (2008). Évaluer des compétences - Guide pratique, Bruxelles : De Boeck.
- Hébert, M.-H., Valois, P. & Frenette, E., (2008), La validation d'outils alternatifs d'évaluation des apprentissages : progressiste ou rétrograde? *Actes du 20e colloque de l'ADMEE-Europe*, Université de Genève.
- Hulin, C. L., Lissak, R. I. & Drasgow, F. (1982). Recovery of two- and three-parameter logistic item characteristic curves : A Monte Carlo study. *Applied Psychological Measurement*, 6, 249-260.
- Jonnaert, P. (2002). *Compétences et socio-constructivisme*. Bruxelles : De Boeck.
- Laveault, D. et Grégoire, J. (1997). Introduction aux théories des tests en psychologie et en sciences de l'éducation. Bruxelles : De Boeck.
- Mottier Lopez, L & Crahay, M. (2009) (Eds.). *Evaluations en tension. Entre la régulation des apprentissages et le pilotage des systèmes*. Bruxelles : De Boeck.

- Rey, B., Carette, V., Defrance, A. & Kahn, S. (2003). Les compétences à l'école — Apprentissage et évaluation, Bruxelles : De Boeck.
- Roegiers, X. (2005), L'évaluation selon la pédagogie de l'intégration : est-il possible d'évaluer les compétences des élèves ?, in Toualbi-Thaâlibi, K. & Tawil, S. (Dir.), *La Refonte de la pédagogie en Algérie - Défis et enjeux d'une société en mutation*, Alger: UNESCO-ONPS, pp.107-124.
- Roegiers, X. (2007). Des situations pour intégrer les acquis scolaires. Bruxelles: De Boeck.
- Scallon, G. (2004). L'évaluation des apprentissages dans une approche par les compétences. Bruxelles : De Boeck Université.

Christophe Dierendonck et Annick Fagnant
Université du Luxembourg



4. RECENSIONS D'OUVRAGES

Gerard, F.M. & Roegiers, X. (2009). Des manuels scolaires pour apprendre, concevoir, évaluer, utiliser. Bruxelles : De Boeck Université

Conçu comme un outil pratique aidant à la conception et à l'évaluation de manuels scolaires, cet ouvrage intéresse autant les auteurs potentiels que les éditeurs ou décideurs soucieux de produire des livres de qualité, les chercheurs en didactique et les enseignants.

Un cadre conceptuel fournit d'abord une assise théorique solide aux processus de conception et d'évaluation des manuels. Une analyse des mécanismes pédagogiques fondamentaux permettant notamment de construire des compétences complète cette approche, débordant ainsi largement du cadre strict induit par les manuels scolaires. Les rapports que ceux-ci peuvent entretenir avec les TICE sont également abordés.

Les auteurs proposent ensuite vingt-huit fiches qui, agrémentées de nombreux exemples et suggestions, permettent au lecteur d'acquérir les compétences nécessaires au rôle de concepteur ou d'évaluateur, voire d'utilisateur de manuels scolaires.

Le point de vue ainsi adopté permet de considérer le manuel tant sous ses aspects pédagogiques, comme outil d'aide à l'apprentissage des compétences, que sous son aspect matériel

Colardyn, D. (2009). *Evaluations et examens aux Etats-Unis, Quelles leçons pour l'éducation en Europe ?* Bruxelles : De Boeck Université

Aux États-Unis, il n'y a pas d'examen national en fin de l'enseignement secondaire. Dans ce pays sans diplôme national, de nombreuses évaluations nationales prolifèrent. Ce livre en présente plusieurs et explique comment elles mesurent les progrès des élèves et des écoles vers l'excellence. Réalisées depuis plus d'un quart de siècle, ces évaluations donnent une image des performances actuelles comme de l'impact des réformes éducatives passées.

La présentation est construite autour de huit questions. Quelles sont les caractéristiques des évaluations ? Quelle est l'efficacité interne de l'éducation et de la formation professionnelle à la *High School* ? Quelle est son efficacité externe ? Les évaluations mesurent-elles le succès des *High School* et témoignent-elles des progrès des élèves dans des domaines fondamentaux comme les mathématiques, la lecture et l'histoire. Les examens avancés et les tests d'entrée dans l'enseignement supérieur illustrent le chemin parcouru comme les progrès à faire. L'épilogue s'interroge sur l'usage des évaluations pour le pilotage des réformes. Comment utiliser les informations issues des évaluations pour concevoir des politiques éducatives plus favorables à l'école pour tous ? En Europe, ces questions sont aussi d'actualité.

La présentation des évaluations et de leurs rôles est illustrée par de nombreux tableaux et graphiques. Des lectures complémentaires et des sites spécifiques sont proposés et un glossaire est inclus.

Les questions traitées peuvent intéresser les responsables éducatifs, économiques et politiques, les chercheurs, les enseignants et les parents. Une meilleure connaissance des questions de « la mesure » dans l'éducation peut éclairer nos systèmes en Europe et nous aider à anticiper certains enjeux.

Romainville, M. & Coggi, C. (Ed.) (2009). *L'évaluation de l'enseignement par les étudiants, Approches critiques et pratiques innovantes*. Bruxelles : De Boeck Université

L'Évaluation de l'Enseignement par les Étudiants (EEE) a progressivement acquis droit de cité dans les universités, malgré les nombreuses résistances qui se sont manifestées à son égard. La situation actuelle de ce type d'évaluation est cependant paradoxale à plus d'un titre. En effet, on dispose désormais d'études nombreuses qui, tout à la fois, confirment la validité et la pertinence de cette forme d'évaluation, mais en pointent aussi les limites, ces dernières rendant notamment nécessaire le recours à d'autres sources complémentaires d'évaluation de l'enseignement. Ces mêmes études identifient les balises à apporter à l'EEE pour la rendre efficace, mais aussi conforme aux valeurs universitaires : respect des spécificités disciplinaires et des contextes, nécessité de suivi ultérieur, ... Or, un inventaire succinct des pratiques les plus courantes d'évaluation de l'enseignement par les étudiants montre que ces précautions et balises, pourtant largement documentées et admises, sont en définitive peu mises en œuvre au sein des dispositifs réels. Ceux-ci ont même tendance à se rabattre sur des formes stéréotypées d'EEE, voire à se limiter à des passations purement administratives de questionnaires standardisés sans que personne ne se soucie ni de leurs effets ni de leurs suivis.

C'est autour de cette contradiction apparente et de cet étrange écart entre la richesse des réflexions et la stéréotypie des pratiques que s'est construit un séminaire de travail qui a rassemblé quatorze spécialistes de l'évaluation de l'enseignement par les étudiants. Le présent ouvrage est le fruit à la fois des présentations qui ont été faites par chaque participant à ce séminaire et des discussions collectives qu'elles n'ont pas manqué de provoquer.



5. ANNONCE DE COLLOQUES ET SEMINAIRES

Congrès AREF 2010 (Actualité de la recherche en éducation et en formation)

du 13 au 16 septembre 2010

à l'université de Genève

Site : <http://www.unige.ch/aref2010/index.html>

5^{ème} réunion du EARLI Special Interest Group « Assessment » :

« Assessment for learners »

du 1^{er} au 3 septembre 2010

au Slaley Hall, Hexham, Northumberland,

NE47 0BX

Site : http://www.northumbria.ac.uk/cetl_afl/earli2010/

2^{ème} réunion du EARLI Special Interest Group « Educational Effectiveness » :

« Efficacité éducative : Modèles, méthodes et Applications »

du 25 au 27 août 2010

au centre pour l'efficacité éducative, Louvain (Belgique)

Site : <http://ppw.kuleuven.be/sigee2010/>

Autres EARLI SIG : http://www.earli.org/special_interest_groups

Congrès ECER 2010 (European Educational Research Association)

« Education and Cultural Change »

Du 25 au 27 août 2010

(pré-conférence : 23-24 août)

Helsinki

Site : <http://www.eera-ecer.eu/ecer/ecer2010/>**6. CONSEIL D'ADMINISTRATION DE L'ADMEE-EUROPE****(issu de l'Assemblée générale du 22 janvier 2009)****Bureau**

Présidente	Lucie Mottier Lopez	Lucie.Mottier@unige.ch
Vice-présidente	Anne Jorro	jorro@univ-tlse2.fr
Secrétaire-trésorier	Vincent Dupriez	Vincent.Dupriez@uclouvain.be

Délégations nationales

Belgique	Vincent Carette	vcarette@ulb.ac.be
	Antoine Derobertmeasure	antoine.derobertmeasure@umh.ac.be
France	Solange Ramond	solange.ramond@orange.fr
	Marc Vantourout	marc.vantourout@parisdescartes.fr
	Nathalie Younes	Nathalie.YOUNES@univ-bpclermont.fr
Luxembourg	Christophe Dierendonck	christophe.dierendonck@uni.lu
	Réginal Burton	reginald.burton@uni.lu
Maroc	Mohammed Talbi	maarifcentre@yahoo.fr
	Hamadi Akrim	hamadi.akrim@gmail.com
Portugal	André Machado	eacm.estp@gmail.com
	Pedro Rodrigues	pedro.rodrigues@fpce.ul.pt
Suisse	Pierre-François Coen	coenp@eduf.fr.ch
	Walther Tessaro	Walther.Tessaro@unige.ch

Autres membres du CA

ADMEE-Internationale	Gérard Figari	gerard.figari@wanadoo.fr
Chargé de mission au secrétariat	Matthis Behrens	Matthis.Behrens@irdp.ch
Rédacteur de la revue MEE	Claude Houssemand	claudio.houssemand@uni.lu
Rédactrice du bulletin	Nathalie Droyer	n.droyer@enesad.fr
Organisatrice du 21 ^{ème} colloque	Catherine Van Nieuwenhoven	contact@catvan.be
Organisatrice du 22 ^{ème} colloque	Palmira Alves	palves@iep.uminho.pt