

L'implantation de l'évaluation assistée par ordinateur (EAO) au sein du système scolaire québécois : avantages et obstacles à surmonter

Éric Dionne

Volume 29, Number 2, 2006

URI: <https://id.erudit.org/iderudit/1086730ar>

DOI: <https://doi.org/10.7202/1086730ar>

[See table of contents](#)

Publisher(s)

ADMEE-Canada - Université Laval

ISSN

0823-3993 (print)

2368-2000 (digital)

[Explore this journal](#)

Cite this article

Dionne, É. (2006). L'implantation de l'évaluation assistée par ordinateur (EAO) au sein du système scolaire québécois : avantages et obstacles à surmonter. *Mesure et évaluation en éducation*, 29(2), 99–116.
<https://doi.org/10.7202/1086730ar>

Article abstract

The computer-assisted assessment (CAA) should take a more and more important place in educational field during the next years. This article adopts a bias in favor of the CAA while exposing the principal advantages and putting emphasis on only certain obstacles to the implementation of CAA in the context of the secondary teaching in Quebec. This article presents a critical synthesis of computer-assisted assessment highlighting the seven following themes: the assessment's moment, the rating, the item banks and the item selection, items and their treatment, the psychological factors linked to CAA, follow-up on the students and the associated costs.

L'implantation de l'évaluation assistée par ordinateur (EAO) au sein du système scolaire québécois : avantages et obstacles à surmonter

Éric Dionne

Université de Montréal

MOTS CLÉS : Évaluation assistée par ordinateur, évaluation des apprentissages

L'évaluation assistée par ordinateur (EAO) est appelée à prendre une place de plus en plus importante en éducation au cours des prochaines années. Cet article adopte un parti pris en faveur de l'EAO en exposant les principaux avantages et en faisant ressortir certains obstacles à la mise en place de l'EAO dans le contexte de l'enseignement secondaire au Québec. Cette démonstration s'effectue en développant les sept thèmes qui suivent, et ce, sous un angle critique, à savoir : le moment de l'évaluation, la correction, les banques d'items et la sélection des items, les items et leur traitement, les facteurs psychologiques liés à l'EAO, le suivi des élèves et les coûts associés.

KEY WORDS : Computer-assisted assessment, evaluation of learning

The computer-assisted assessment (CAA) should take a more and more important place in educational field during the next years. This article adopts a bias in favor of the CAA while exposing the principal advantages and putting emphasis on only certain obstacles to the implementation of CAA in the context of the secondary teaching in Quebec. This article presents a critical synthesis of computer-assisted assessment highlighting the seven following themes : the assessment's moment, the rating, the item banks and the item selection, items and their treatment, the psychological factors linked to CAA, follow-up on the students and the associated costs.

Note de l'auteur : L'auteur tient à remercier monsieur Jean-Guy Blais, Ph.D., professeur titulaire en mesure et évaluation à l'Université de Montréal, ainsi que les trois arbitres anonymes pour leurs commentaires fort pertinents qui ont permis de bonifier cet article. Merci également à madame Catherine Bertrand pour avoir assuré la révision linguistique de ce texte. Toute correspondance peut être adressée par courriel à l'adresse suivante : [eric.dionne@umontreal.ca].

PALAVRAS-CHAVE: Avaliação assistida por computador, avaliação das aprendizagens

A avaliação assistida por computador (AAC) será chamada a desempenhar um papel cada vez mais importante na educação, nos próximos anos. Este artigo adopta uma posição favorável à AAC, expondo as principais vantagens e ressaltando alguns obstáculos à implementação da AAC no contexto do ensino secundário no Québec. Esta demonstração efectua-se pelo desenvolvimento de sete temas a partir de uma abordagem crítica, a saber: o momento da avaliação, a correcção, os bancos de itens e a selecção dos itens, os itens e o seu tratamento, os factores psicológicos ligados à AAC, o acompanhamento dos alunos e os custos associados.

Introduction

Au cours des dernières décennies, les progrès réalisés en informatique ont été particulièrement remarquables. Ces derniers ont permis de mettre sur le marché des ordinateurs et des serveurs de plus en plus rapides et performants. L'évolution dans ce domaine a aussi permis d'offrir une kyrielle de périphériques (micros, caméras numériques, appareils photos numériques, sondes, etc.) qui permettent de recueillir une information qu'il serait plus difficile à obtenir avec des outils de type papier-crayon. Il est maintenant possible de traiter des données de divers ordres (cognitif, moteur, affectif, etc.) et de différents champs (psychologie, éducation, gestion du personnel en entreprise, etc.) avec de plus en plus de rapidité et d'efficacité. Jumelé à ceci, notons la baisse des coûts d'achat et d'entretien associés à ces appareils, qui les rend ainsi de plus en plus abordables, notamment pour le secteur de l'éducation (Bennett, 2002a).

Les technologies de l'information et de la communication offrent des perspectives intéressantes, notamment en évaluation des apprentissages en contexte scolaire. Pour les besoins de notre analyse, nous avons choisi de limiter l'évaluation assistée par ordinateur (EAO) au seul contexte de l'enseignement secondaire (évaluation à petite échelle, c'est-à-dire à l'échelle de la classe) ce qui représente, somme toute, un champ important et florissant de l'EAO pour le Québec.

Le présent article adopte un parti pris favorable en exposant les principaux avantages et en ne faisant ressortir que certains obstacles à la mise en place de l'EAO dans le contexte de l'enseignement secondaire au Québec. Pour ce faire, nous développerons les sept thèmes qui suivent, et ce, sous un angle critique, à savoir: le moment de l'évaluation, la correction, les banques d'items et la sélection des items, les items et leur traitement, les facteurs psychologiques

liés à l'EAO, le suivi des élèves et les coûts associés. Le choix de ces thèmes repose sur le fait qu'ils supportent le propos actuel en plus d'être fréquemment utilisés par les auteurs dans le domaine (Alderson, 2000; Blais, Dionne & Raïche, 2005; Brown, Race, & Bull, 1999; Drasgow & Olson-Buchanan, 1999a; Luecht, 2006). Deux petites observations avant d'aborder le cœur du sujet. D'abord, le lecteur¹ remarquera que plusieurs aspects, par exemple les risques de tricherie, sont présents dans plus d'un thème. Ceci s'explique par le fait que la plupart des aspects traités sont liés entre eux et peuvent ainsi difficilement être isolés des autres aspects traités dans d'autres thèmes. Ensuite, nous aurons recours, lorsque cela sera pertinent à notre propos, à des exemples tirés de certains modèles plus spécifiques d'évaluation assistée par ordinateur, comme les tests adaptatifs.

L'évaluation assistée par ordinateur : définition et contexte

Une définition de l'EAO

Bien que l'on retrouve une documentation scientifique relativement abondante sur l'EAO, il est étonnant de constater que bien peu d'auteurs s'aventurent à définir l'expression «évaluation assistée par ordinateur». Dans ce texte, nous nous baserons sur les écrits de Bunderson, Inouye et Olsen (1993) qui considèrent que l'évaluation en éducation est un processus constitué de trois sous-processus :

- le développement du test (définir les caractéristiques du test, procéder au prétest, réviser les items, etc.);
- l'administration du test (obtenir les réponses, établir un score, interpréter le score, etc.);
- l'analyse du test (établir l'équivalence, valider, rechercher le fonctionnement différentiel d'item [FDI], etc.).

Ainsi, l'évaluation assistée par ordinateur peut être définie par le recours à des outils informatiques permettant d'optimiser ou de faciliter la réalisation de chacun de ces sous-processus.

Les technologies de l'information et des communications permettent d'offrir un contexte d'apprentissage et d'évaluation relativement différent du contexte standard papier-crayon. À titre d'exemple, elles permettent de grandement varier le format des items et des tests. On retrouve dans la documentation plusieurs exemples d'utilisations des technologies aux fins d'évaluation des apprentissages. Par exemple, le *Portland Public School* aux États-Unis admi-

nistre des tests adaptatifs² s'appuyant sur la théorie des réponses aux items (TRI) dans cinq domaines disciplinaires (Kingsbury & Houser, 1999). Toujours aux États-Unis, quatre États (Dakota, Oregon, Georgie, Virginie) ont commencé à progressivement fournir des outils d'évaluation via le Web³ (Bennett, 2002b). Aussi, une multitude d'items statiques (sans interactivité avec le sujet) et dynamiques (où le sujet doit réaliser des choix) sont maintenant mis au point en lien avec ces nouvelles technologies. Nous reviendrons plus en détail sur cet aspect quand nous aborderons le sujet des items et de leur traitement. En somme, l'EAO offre des outils de plus en plus sophistiqués permettant de planifier, réaliser et réguler une démarche d'évaluation.

Le contexte pédagogique dans lequel on situe l'EAO

Le contexte entourant l'évaluation des apprentissages peut être passablement vaste. On peut faire référence aux résultats obtenus par un élève, un groupe-classe, une école, une commission scolaire ou l'ensemble des élèves d'une province, d'un État ou d'un pays. Bien que la documentation fasse principalement mention de l'EAO dans un contexte d'évaluation à grande échelle⁴, il nous est apparu plus pertinent de la situer – dans ce texte – dans une perspective plus restreinte à savoir à l'échelle de la classe (classroom assessment). Dans cet article, nous nous baserons sur la définition de McMillan (2001, p. 9) qui considère l'évaluation à l'échelle de la classe comme «la collection, l'évaluation et l'utilisation de l'information afin d'aider l'enseignant à prendre les meilleures décisions possibles». Nous avons retenu ce contexte pour deux raisons principales. D'abord, les initiatives de recherche québécoise concernant l'EAO se situent rarement à grande échelle. Ensuite, parce que les recherches et les innovations concernent souvent d'abord l'évaluation formative des apprentissages, qui implique des enjeux moins importants pour les différents intervenants (élèves, parents, enseignants, etc.) que, par exemple, l'évaluation certificative. En résumé, dans ce texte, à moins de l'indiquer, nous aborderons les différents aspects de l'EAO dans une perspective d'évaluation formative à l'échelle de la classe.

Le moment de l'évaluation

L'évaluation en contexte scolaire se caractérise généralement par le fait que tous les élèves⁵ subissent la même épreuve sous format papier-crayon au même moment et dans les mêmes conditions d'administration. Au secondaire, tous les sujets se voient généralement offrir le même examen (les mêmes items) à une même période donnée de l'horaire dans un seul et même local et avec les mêmes consignes. Cette description correspond à ce que Scallan

appelle une «standardisation des procédés dans une perspective traditionnelle» (Scallon, 2004, p. 24). Cette description a notamment été constatée sur le terrain par Dionne (1999) qui a alors procédé à l'analyse des épreuves que les enseignants de sciences d'une commission scolaire québécoise utilisaient en cours d'année scolaire pour rendre compte des apprentissages de leurs élèves. Cette uniformisation des conditions de passation des épreuves évaluatives a l'inconvénient de ne pas tenir compte des différences intra-groupes (Drasgow & Olson-Buchanan, 1999b). Ainsi, un sujet possédant un rythme d'apprentissage plus rapide est freiné dans sa séquence d'apprentissage alors qu'un sujet plus lent pourrait avoir besoin de plus de temps pour accomplir les mêmes apprentissages qu'un sujet plus rapide. L'EAO permet de contrer cet écueil puisque les sujets peuvent choisir, avec l'accord de l'enseignant, le moment où ils sont prêts à subir l'évaluation (Sands, Waters & McBride, 1997). Un contexte d'EAO peut aider les enseignants à différencier leur pédagogie en tenant davantage compte du rythme d'apprentissage et des différences individuelles de chacun des élèves. Par exemple, certains sujets particulièrement doués ou rapides pourraient se voir offrir des défis plus substantiels après avoir rendu compte de leurs apprentissages au moyen d'un test informatisé qu'ils auraient subi avant les autres élèves de la classe, l'EAO offrant un contexte d'administration plus souple par rapport aux tests papier-crayon. En effet, l'EAO permet une forme d'évaluation continue où chaque sujet est évalué au moment où il se sent en mesure de faire montre de ses apprentissages (Carlson, 1993). À titre d'exemple, cette situation peut être maximisée à l'aide de tests adaptatifs, c'est-à-dire lorsque le choix de l'item à administrer est fonction de la compétence de chaque élève, ce qui se manifeste par le vecteur-réponse dudit élève. Dans ce cas, non seulement celui-ci peut répondre à son propre rythme mais le temps de passation (indirectement le nombre d'items) peut être beaucoup plus court qu'en contexte d'évaluation papier-crayon traditionnelle (English, Reckase & Patience, 1977; Olsen et al., 1989; Wainer et al., 2000). Cette forme d'EAO permet d'ajuster l'épreuve afin de fournir des items qui permettent de mesurer plus précisément les apprentissages ou le niveau de compétence de l'élève. Par contre, l'évaluation continue pose également un certain nombre de problèmes. Sur le plan technique, elle pose celui d'offrir un environnement adéquat (p. ex., silencieux) aux personnes qui subissent une épreuve alors que les autres élèves sont en apprentissage dans la classe (Kingsbury & Houser, 1999). En effet, plusieurs élèves ont besoin, et avec raison, d'un environnement calme et silencieux afin de se concentrer sur leur tâche. Se pose également le problème lié à la sécurité qui peut s'actualiser,

notamment, par les risques de tricherie. Un environnement informatique, tout comme un environnement papier-crayon du reste, ne garantit pas que les élèves ne chercheront pas à consulter leurs documents ou leurs pairs alors que l'épreuve est, par exemple, individuelle et à livre fermé. En ce sens, l'individualisation des évaluations demeure une pratique qui nécessite une gestion serrée de l'environnement dans lequel l'évaluation se réalise.

La correction

Un des avantages manifestes de l'EAO réside dans le temps économisé lors de la correction des épreuves. En effet, un environnement informatique permet de réaliser la correction et la compilation des résultats en un temps minimal. Les enseignants n'ont donc pas à passer de longues heures à corriger des copies et à comptabiliser des scores puisque l'ordinateur s'en occupe. Présentement, cela est possible avec différents types d'items à correction objective (items à choix multiples, items à appariement, etc.). Cependant, on peut penser que dans un proche avenir il sera possible de traiter des items à réponses élaborées grâce aux développements issus de la recherche en correction automatique de rédactions écrites (*automated essay scoring*) (Shermis & Burnstein, 2003). Par contre, comme le souligne Alderson (2000), il reste encore beaucoup à faire avant d'être en mesure de proposer des items à développement long qui pourront être par la suite traités de façon satisfaisante.

Avec l'informatique, la correction est non seulement plus rapide mais elle est également plus sûre puisqu'elle est réalisée numériquement et non mécaniquement ou manuellement (Russo, 2002; Sands et al., 1997). À titre d'exemple, plusieurs enseignants d'écoles secondaires québécoises utilisent des procédés mécanographiques pour la correction de tests bien que la plupart de ceux-ci soient corrigés manuellement. Ces façons de faire ne sont pas exemptes d'erreurs de nature humaine ou mécanique. Le traitement numérique des copies d'élève permet donc de minimiser les risques d'erreurs liées directement aux correcteurs (erreur de calcul, erreur d'inattention, etc.) et celles liées aux manipulations des feuilles réponses (Wainer et al., 2000) qui sont alors, rappelons-le, éliminées.

Puisque la correction se fait par ordinateur, il est possible d'obtenir les résultats de l'évaluation dès la fin de la passation. On peut ainsi cibler facilement les items qui ont été ratés ou réussis. Ceci est généralement très apprécié par les élèves qui subissent l'épreuve (Epstein et al., 2002; Glowacki, McFadden & Price, 1995; Laurier, 1999; Wainer et al., 2000) puisqu'ils ont la possibilité d'obtenir une rétroaction rapide à la suite de leur performance. Dans un

contexte d'évaluation en cours d'apprentissage, ceci est un avantage manifeste (Trentin, 1997). Le support informatique permet ainsi de cerner plus facilement les difficultés des élèves et d'apporter les ajustements nécessaires. La révision d'un examen est souvent laborieuse dans les classes où les élèves subissent des examens de type papier-crayon. Tous doivent écouter au rythme des explications de l'enseignant même si les questions auxquelles il répond ont été réussies par l'élève. En individualisant l'approche, cela permet à chaque élève d'apporter les correctifs qui s'appliquent à son propre cas. Pour l'enseignant ou le professeur, les technologies de l'information et de la communication permettent de gérer de façon plus efficace l'information devant faire l'objet de la rétroaction.

Les banques d'items et la sélection des items

Plusieurs problèmes liés à la tricherie sont accolés à l'évaluation de type papier-crayon : vol de copies d'examen, bouche à oreille sur les questions les plus fréquemment posées, etc. L'EAO permet de limiter ces inconvénients (Sands et al., 1997; Wainer et al., 2000) puisque les items sont puisés dans des banques contenant généralement un nombre important d'items. Étant en grand nombre, il est plus difficile de mémoriser ces derniers afin de les divulguer à une autre personne. Aussi, puisqu'il n'y a ni questionnaire, ni copie d'examen, les risques de vol ou de perte sont minimisés. Certains auteurs (Davey & Nering, 2002; Schnipke & Scrams, 1999) sont moins optimistes face aux risques de tricherie. Ils rappellent d'ailleurs les problèmes liés au « cas » GRE (*Graduate Record Examination*) survenu en 1994, qui avait montré les failles dans la sécurité d'un test adaptatif offert par l'organisme états-uniens ETS (*Educational Testing Service*). À ce moment, des individus avaient noté les items les plus fréquemment administrés. Cet événement a rappelé l'importance et la difficulté d'assurer la sécurité des banques d'items afin de ne pas compromettre les qualités métrologiques des items présents dans les banques. À cet égard, trois éléments importants sont à considérer pour réduire les risques de tricherie : minimiser les risques de plagiat entre deux ou plusieurs élèves (par exemple quand un élève copie intégralement la réponse d'un autre élève en regardant ce que ce dernier a fourni comme réponse sur l'écran cathodique), s'assurer de l'identité de l'élève qui réalise l'épreuve et réduire les risques que des items soient à la disposition des élèves avant l'évaluation (Davey & Nering, 2002). Ainsi, bien que l'EAO minimise certains risques de tricherie, en contrepartie elle s'expose à d'autres types de tricherie, comme l'intrusion de pirates informatiques dans les banques d'items.

Un autre problème important associé aux banques d'items réside en la construction de ces dernières et en leur entretien. Bien que ces aspects ne soient pas exclusifs au traitement «informatisé» des items, ils n'en demeurent pas moins une source importante de préoccupation. Lors de la construction d'une banque d'items, Anzaldúa (2002) suggère de tenir compte des cinq éléments qui suivent :

- les items de la banque doivent mesurer un construit utile ;
- les items doivent être organisés ;
- les items doivent être suffisamment nombreux ;
- les items doivent être expérimentés pour s'assurer de leur caractère unidimensionnel⁶ ;
- les items doivent être calibrés.

Les éléments précédents nous permettent de constater qu'une banque d'items est beaucoup plus qu'un simple amalgame d'items. Celle-ci doit être bâtie afin de tenir compte de plusieurs aspects (modèle d'EAO utilisé, importance du test, niveau de sécurité souhaité, etc.) et remplir ainsi plusieurs rôles névralgiques en évaluation assistée par ordinateur.

Plusieurs modèles d'EAO nécessitent des banques d'items contenant plusieurs centaines d'items différents. Anzaldúa (2002) suggère d'ailleurs de centraliser au moins trois fois le nombre d'items nécessaire dans la banque. Cette façon de procéder permet de s'assurer d'avoir un nombre suffisant d'items pour créer des tests différents. Cependant, cette contrainte pose le défi de rédiger un nombre important d'items qui soient valides (Glowacki et al., 1995).

Une banque d'items a continuellement besoin d'être entretenue et mise à jour. Avec le temps, on doit en ajouter pour combler le retrait de certains items qui ne sont plus adéquats. On doit également en réviser un certain nombre afin de les ajuster en fonction de leurs qualités métrologiques. En effet, un environnement d'évaluation informatisé permet de prétester facilement des items (Sands et al., 1997 ; Wainer et al., 2000) et de recueillir ainsi des données permettant de prendre des décisions à leur sujet (les éliminer, les ajuster, etc.).

La sélection des items est également une opération complexe en EAO. Pour ce faire, il faut élaborer un algorithme de sélection d'items qui tient compte, notamment, des variables suivantes :

- les caractéristiques du modèle d'EAO choisi ;
- le niveau d'exposition des items ;
- l'essai de nouveaux items.

Il existe plusieurs modèles d'EAO. Parmi ceux-ci, on peut citer le WBT (*Web-based testing*), les TA (tests adaptatifs) ou le CAST (*Computerized-adaptive sequential testing*). Le choix du modèle aura un impact important sur l'architecture de l'algorithme de sélection d'items. Par exemple, un test adaptatif de type CAST s'appuyant sur le modèle de Rasch devra tenir compte de l'information (I) mais aussi de la représentativité du domaine disciplinaire évalué. Par contre, un modèle simple de tests informatisés qui ne fait que présenter des tests préassemblés n'a pas besoin d'un algorithme de sélection d'items très sophistiqué puisque seule la représentativité du contenu disciplinaire prime alors.

En raison de contraintes de sécurité, un algorithme de sélection d'items doit permettre de tenir compte du niveau d'exposition des items. Il faut éviter que ce soit continuellement les mêmes items qui soient constamment administrés aux élèves. À cet égard, plusieurs stratégies sont disponibles. Examinons deux de ces stratégies utilisées fréquemment dans le cadre de tests adaptatifs. La première méthode est celle élaborée par McBride et Martin (1983). Cette méthode consiste à distinguer les quatre items fournissant une information (I) semblable et de les ordonner ensuite sous la forme 4-3-2-1. L'item 1 est utilisé dans 10% des cas, l'item 2 dans 20% des cas, l'item 3 dans 30% des cas et l'item 4 dans 40% des cas (Davey & Nering, 2002). C'est donc une stratégie probabiliste dont le choix de l'item relève du hasard. La seconde méthode est celle de Sympson et Hetter (1985). Elle consiste à assigner à chacun des items un paramètre d'exposition compris entre 0 et 1. Ainsi, un item ayant un paramètre d'exposition de 0,5 sera administré une fois sur deux lorsqu'il sera sélectionné. Cette stratégie vise à combler les lacunes de la méthode de McBride et Martin, qui est moins efficace pour les items choisis fréquemment et pour ceux choisis très rarement (Davey & Nering, 2002).

Enfin, l'algorithme de sélection d'items doit tenir compte des besoins liés aux essais de nouveaux items. Comme nous l'avons vu dans une section antérieure, il faut constamment faire la mise à jour d'une banque d'items en enlevant, modifiant ou ajoutant des items. Ainsi, on peut déterminer un pourcentage d'items qui seront expérimentés lors des tests. Ce pourcentage sera fonction du nombre d'items présents dans la banque, des besoins d'expérimentation, etc.

Les items et leur traitement

En contexte d'évaluation papier-crayon, il est peu fréquent – dans la pratique – d'analyser les réponses des élèves. Plusieurs hypothèses peuvent être avancées pour justifier ceci : le temps nécessaire pour effectuer de telles analyses, l'absence d'outils utiles et efficaces, la compétence déficiente des évaluateurs, etc. On se contente généralement de comptabiliser les résultats et de remettre un résultat global au moyen d'un score exprimé souvent en pourcentage. L'analyse des items est donc rare et les élèves ont peu d'occasions de prendre note de leurs forces et de leurs faiblesses, ce qui est particulièrement ennuyeux dans une perspective d'aide à l'apprentissage. En contexte d'ÉAO, il devient possible d'analyser les items beaucoup plus souvent car le traitement est fait par l'ordinateur en un temps relativement court. Aussi, le traitement informatique des items et des réponses permet de recueillir davantage de données qu'avec un mode d'administration papier-crayon. Ce dernier permet généralement de traiter uniquement la réponse de l'élève. Avec l'évaluation assistée par ordinateur, il est aussi possible de colliger beaucoup plus de données. On peut penser, par exemple, au temps de réponse des élèves à chaque item (Wainer et al., 2000). Cette donnée peut permettre, notamment, de donner des indices permettant de retracer les élèves qui, par exemple, répondent au hasard.

L'évaluation assistée par ordinateur offre la possibilité d'administrer des items fort différents de ceux normalement administrés sous une forme papier-crayon. Plusieurs auteurs (Bennett, 2002b; Bunderson et al., 1988; Bunderson et al., 1993; Drasgow & Mattern, 2006; Drasgow & Olson-Buchanan, 1999b; Mills, Potenza, Fremer & Ward, 2002; Natal, 1998; Rabinowitz & Brandt, 2001; Wainer et al., 2000) présentent des exemples de « nouveaux » items que l'informatique permet maintenant de créer et d'utiliser. En science, mentionnons les simulations informatiques qui permettent de plonger l'élève dans une situation d'évaluation authentique, où il doit résoudre un problème de physique dans un laboratoire virtuel. Dans cet exemple, on voit qu'une forme d'évaluation papier-crayon (situation statique en deux dimensions) ne permet pas un contexte aussi riche qu'en version électronique (situation dynamique). En formation personnelle et sociale, il est possible d'avoir recours à des extraits vidéo afin de juger de certains comportements ou de certaines attitudes. Dans ce cas, l'informatique permet de présenter une situation plus riche (offrant davantage de subtilités) qu'elle ne le serait sur papier. Cet exemple montre, une fois encore, les possibilités qu'offre l'informatique quant à la rédaction de nouvelles situations d'évaluation et de nouveaux items. Cependant, la validité

de ces items reste, dans bien des cas, à démontrer. Certains auteurs (Rosenquist, Shavelson & Ruiz-Primo, 2000) remettent d'ailleurs en question la validité des items électroniques versus les items papier-crayon mentionnant, notamment, qu'il n'est pas simple de démontrer l'équivalence de la validité entre les formes électroniques et papier-crayon d'items semblables. Autrement dit, il est difficile de démontrer que les deux formes mesurent le même construit. Une condition essentielle à l'utilisation et à l'acceptation des tests informatisés est de montrer leur équivalence à un test similaire sous format papier-crayon. Cependant, cette démonstration n'est pas sans soulever plusieurs difficultés (Wainer et al., 2000). L'étude de Neuman et Baydoun (1998) indique qu'en suivant certaines conditions, il n'existe que peu de différences entre les versions électroniques et papier d'un test sur le plan psychométrique. Il en est de même avec la recherche de Olsen (1989) qui conclut que les résultats sont identiques mais accessibles en moins de temps en utilisant un test adaptatif. En somme, le problème de l'équivalence des items papier-crayon et informatisés n'est pas réglé et de plus amples recherches restent à faire sur cette question.

Certaines habiletés cognitives (planification, estimation, etc.) semblent difficiles à évaluer par le biais de l'informatique (Trentin, 1997). Il en est de même pour certaines compétences nouvellement prescrites dans les récents programmes maintenant en vigueur au premier cycle du secondaire (Ministère de l'Éducation du Québec, 2003). Par exemple, la plupart des compétences transversales⁷ peuvent difficilement être évaluées par le biais d'items traditionnels sous une forme papier-crayon ou informatisée. Dans le contexte scolaire québécois actuel, on constate que les tests informatisés qui utilisent des items à correction objective ne semblent pas être les plus appropriés pour évaluer des compétences. En ce sens, ces tests informatisés pourraient être utilisés, notamment, dans une perspective d'évaluation formative ayant comme objet des éléments disciplinaires spécifiques. Examinons un exemple pouvant s'appliquer à la science et à la technologie. Dans ce contexte, il pourrait s'agir de déterminer si l'élève est en mesure de bien représenter ses données – issues d'une expérimentation qu'il aura conçue et réalisée – sous une forme graphique. Dans cet exemple, l'élève ne pourrait faire montre de l'ensemble de sa démarche mais il pourrait démontrer certaines habiletés ou techniques dans le cadre d'une EAO. Cette dernière pourrait être utile pour recueillir des données sur certains aspects de la démarche. On peut aussi penser que les tests informatisés pourraient également être adaptés pour évaluer des processus cognitifs un peu plus complexes. Toujours en science et technologie, l'expérimentation assistée par ordinateur (EXAO) permet déjà de concevoir,

de réaliser et d'évaluer des expériences. L'EXAO permet de recueillir des données, de les traiter en vue de tirer des conclusions sur certaines hypothèses en lien avec un phénomène scientifique (Nonnon, 2005). Ce genre de dispositif peut ainsi permettre de traiter certains phénomènes qui permettent de rendre compte du niveau de compétence d'un élève en science et technologie.

Le traitement numérique des items permet également de bâtir des banques d'items qui offrent la possibilité de centraliser les items disponibles sur un élève donné. Ceci est une perspective intéressante qui évite un dédoublement d'énergie et de temps en permettant aux enseignants de partager des items qui auront été expérimentés par de nombreux élèves. La construction de banques d'items volumineuses se trouvent ainsi justifiée, de même que leur entretien, en raison du nombre de personnes que celles-ci peuvent desservir (Dragow & Olson-Buchanan, 1999b). En définitive, il s'agit de réaliser une mise en commun des items et des tests afin de maximiser les différents sous-processus d'une démarche d'évaluation en vue de la bonifier.

Enfin, précisons qu'au moment de la passation d'une épreuve, un environnement informatique offre de plus grandes possibilités d'ajustements en fonction de besoins particuliers que peuvent nécessiter certaines personnes ou certains groupes de personnes. (Glowacki et al., 1995; Wainer et al., 2000). À titre d'exemple, un environnement informatique permet d'administrer des items dictés par l'ordinateur, ce qui peut accommoder les sujets ayant des difficultés visuelles. En ce sens, le recours à l'informatique peut assurer une meilleure intégration d'élèves ayant des besoins particuliers en leur permettant de réaliser les évaluations au moment qui leur convient, comme tous les autres élèves. Avec la volonté politique de plus en plus manifeste du ministère de l'Éducation, du Loisir et du Sport du Québec (MELS) de favoriser l'intégration des personnes ayant des difficultés physiques ou intellectuelles, on peut penser qu'un environnement informatique pourrait faciliter cette intégration.

Les facteurs psychologiques liés à l'EAO

Certaines personnes ne sont pas familières avec l'utilisation des ordinateurs et elles peuvent être perturbées de subir un test sur support informatique. Malgré tout, certaines études tendent à démontrer que ce problème est minimal chez les jeunes élèves (Wang & Chuang, 2002). Une autre étude (Powell, 1994) arrive à une conclusion similaire avec des élèves ayant réalisé un test informatisé de type adaptatif. Aussi, l'anxiété face à l'utilisation de l'ordinateur serait, selon certaines études, différente selon que les élèves soient des filles ou des garçons, alors que d'autres études prétendent que l'anxiété peut

être contrée facilement en initiant les élèves très tôt à un environnement informatique. Stephens (2001) rapporte qu'il est possible de diminuer l'anxiété des élèves à l'aide de prétests et d'explications adéquates sur la nature et l'objet de l'évaluation. Harris (2006), ajoute deux actions «pragmatiques» aux suggestions de Stephens permettant d'améliorer la réaction de sujets dans le cadre d'une évaluation via le Web (*internet-based testing*):

1. informer les individus de ce qu'il adviendra des données recueillies et de la façon dont elles seront traitées et emmagasinées,
2. s'assurer qu'une personne (et non seulement une machine) puisse aider un individu qui en manifesterait le besoin.

En somme, bien que les résultats de recherche ne soient pas consensuels sur cette question, on peut quand même noter que le sexe, l'âge et le degré de familiarisation avec les ordinateurs sont trois facteurs cruciaux à considérer dans une perspective d'évaluation assistée par ordinateur.

Le suivi des élèves

Plusieurs enseignants se plaignent de la difficulté qu'ils ont à assurer le suivi des apprentissages de leurs élèves, principalement quand leurs groupes sont nombreux ou quand ils ont plusieurs groupes (ce qui est la norme dans le contexte de l'enseignement secondaire au Québec). Généralement, les pédagogues consignent des résultats chiffrés qui font office de «témoins» des apprentissages. Cependant, une note ne permet pas de refléter, pour chaque individu, l'état de ses apprentissages, à savoir ses forces et ses faiblesses. Les enseignants conservent aussi parfois des évaluations sous forme papier mais la gestion de l'ensemble de ces documents peut devenir rapidement contraignante et difficile. En utilisant les moyens informatiques, il est plus facile, d'un point de vue pédagogique et administratif, de «suivre» le cheminement d'un élève à travers son parcours de formation. Par exemple, il devient plus facile de conserver les traces des apprentissages d'un élève si celui-ci a été évalué au moyen d'outils informatiques. Cette possibilité permet également d'assurer une meilleure régulation auprès de l'élève en travaillant de façon plus ciblée sur les aspects de sa formation qui semblent problématiques au regard de l'évaluation puisqu'il est possible de faire référence à des données issues d'évaluations antérieures.

Le recours à l'EAO peut aussi être avantageux si l'élève change d'école ou de commission scolaire: il est facile de faire suivre son dossier numérique contenant ses évaluations, ses résultats et même ses travaux représentatifs de ses apprentissages sous la forme, par exemple, d'un portfolio électronique.

Ceci peut permettre une transition plus harmonieuse en s'assurant que l'information pédagogique de l'élève l'accompagne avec une faible perte de données. Cependant, cette perspective implique un souci de confidentialité important afin d'éviter que des données soient perdues ou remises entre les mains de personnes qui ne devraient pas avoir accès à cette information.

Les coûts liés à l'EAO

L'évaluation assistée par ordinateur implique des coûts d'implantation et de support logistique particulièrement importants (Sands et al., 1997). Ces derniers sont dus, en partie, aux facteurs qui suivent :

- achat d'équipement (ordinateurs pour les usagers, serveurs, équipements pour le réseau, imprimantes, etc.);
- mise en place et entretien (p. ex., la banque d'items);
- formation des personnes responsables de la passation (enseignants, formateurs, etc.).

Certains auteurs (Davey & Nering, 2002) prétendent que les coûts associés à un test administré par ordinateur sont de deux à trois fois supérieurs à ceux engendrés pour un test papier-crayon. Concernant l'achat du matériel, Natal (1998) mentionne que les coûts liés à l'acquisition de l'équipement informatique sont de moins en moins problématiques. Selon cet auteur, les ordinateurs fabriqués au cours des cinq dernières années peuvent être tout à fait adéquats pour une utilisation impliquant l'EAO. Nonobstant ceci, mentionnons que les facteurs économiques freinent considérablement la percée de l'EAO dans les différents milieux éducatifs. Dans le contexte québécois, la disponibilité des ordinateurs est de moins en moins problématique au primaire et au secondaire. Bien que certaines études (Auger, 1989) démontrent les multiples possibilités d'implantation de modèles d'EAO, les initiatives en ce sens demeurent encore aujourd'hui de l'ordre de l'exception. Les coûts ne sont pas étrangers à cette situation.

En revanche, mentionnons que les tests informatisés permettent de réaliser des économies d'impression puisque les copies d'examen et les feuilles de réponses sont éliminées. On peut également ajouter à ceci les coûts indirects liés à l'impression (personnel, équipement, etc.) bien que la plupart des institutions scolaires ne peuvent se permettre de se passer d'un service d'impression.

Les coûts liés à l'entretien sont également importants dans l'implantation d'une infrastructure d'EAO. Il faut des techniciens qualifiés qui assurent le bon fonctionnement de l'équipement afin d'éviter les pannes, les avaries et tous les autres imprévus pouvant survenir en cours d'utilisation.

Enfin, la formation des enseignants est indispensable pour assurer le fonctionnement harmonieux d'un tel environnement d'évaluation. Cela nécessite non seulement une formation initiale importante mais également une formation continue, qui s'adapte aux innovations qui ont cours dans le domaine des tests informatisés.

Conclusion

L'évaluation assistée par ordinateur est appelée à prendre une place sans cesse grandissante en évaluation des apprentissages. Dans le cadre de cet article, nous avons adopté une position biaisée favorable à l'essor de l'EAO ce qui représente, en soi, une limite à l'analyse que nous avons réalisée. Cependant, nous pensons que l'essor de l'EAO est incontournable puisque les gains dépassent largement les efforts consentis. Parmi les raisons qu'il est possible d'invoquer pour justifier l'expansion de l'EAO, mentionnons :

- la possibilité d'offrir des items ajustés aux compétences de l'élève, et ce, au moment où celui-ci est prêt à faire état de ses habiletés ou de ses compétences ;
- la réduction du temps de préparation des épreuves et du temps de correction ;
- la plus grande facilité à assurer un suivi individualisé à chacun des élèves ;
- la possibilité de rédiger de nouveaux items qui mettent à profit la puissance du traitement de l'information par les ordinateurs ;
- la possibilité de la mise en commun de l'information et de la centration de cette dernière à des fins d'analyse et d'amélioration du processus.

Cependant, plusieurs problèmes risquent de freiner la progression de l'EAO. Ces problèmes nécessitent des recherches plus poussées afin de trouver des pistes de solution permettant de surmonter les obstacles. Parmi les défis qu'il faudra surmonter, mentionnons :

- les coûts liés à la mise en place de l'infrastructure technique ;
- les contraintes actuelles liées au format des items (à correction objective) ;
- la formation des enseignants et des intervenants en éducation sur l'évaluation assistée par ordinateur (ajouter/modifier des items dans une banque d'items, utiliser efficacement l'interface, gérer les résultats des élèves, etc.).

L'évaluation assistée par ordinateur n'est pas encore une pratique courante mais il est raisonnable de projeter qu'elle est appelée à prendre une importance sans cesse croissante au cours des prochaines années. En ce sens, la recherche dans ce domaine est une condition sine qua non à son expansion en milieu scolaire et en particulier à l'ordre d'enseignement secondaire au Québec.

NOTES

1. Le masculin est employé dans le seul but d'alléger ce texte.
2. Dans la documentation on retrouve principalement l'acronyme CAT pour «*computerized adaptive testing*».
3. On fait ici référence au «*Web-based testing*» (WBT).
4. La documentation relative à l'EAO est extrêmement abondante aux États-Unis d'où viennent la plupart des recherches concernant ce champ d'étude.
5. Dans ce texte, nous considérons équivalents les termes «élèves», «étudiants», «sujets» et «apprenants».
6. Principalement dans le cas où on s'appuie sur un modèle de TRI.
7. Quelques exemples de compétences transversales : coopérer, structurer son identité, etc.

RÉFÉRENCES

- Alderson, J. C. (2000). Technology in Testing: The Present and the Future. *System*, 28(4), 593-603.
- Anzaldua, R. M. (2002). *Item Banks: What, Where, Why and How*. Paper presented at the 25^e Annual Meeting of the Southwest Educational Research Association, Austin, TX, February 14-16, 2002.
- Auger, R. (1989). *Étude de praticabilité du testing adaptatif de maîtrise des apprentissages scolaires au Québec*. Thèse de doctorat, Université du Québec à Montréal.
- Bennett, R. E. (2002a). Inexorable and inevitable: The continuing story of technology and assessment. *The journal of technology, learning, and assessment*, 1(1), 1-24.
- Bennett, R. E. (2002b). Using electronic assessment to measure student performance: online testing. *The state education standard*, 3(3), 22-29.
- Blais, J.-G., Dionne, E., & Raïche, G. (2005). *Défis et enjeux de l'évaluation des apprentissages à l'aide des technologies de l'information et de la communication*. Actes du 18^e colloque international de l'ADMEE, Reims, France.
- Brown, S., Race, P., & Bull, J. (1999). *Computer-assisted assessment in higher education*. London: Kogan Page.
- Bunderson, C. V., et al. (1988). *The Four Generations of Computerized Educational Measurement*. Princeton, NJ: Educational Testing Service. (Evaluative/Feasibility ETS-RR-88-35)
- Bunderson, C. V., Inouye, K. D., & Olsen, J. B. (1993). The four generations of computerized educational measurement. In L. R. Linn (Ed.), *Educational measurement* (3^e éd., pp. 367-407). Phoenix: National council on measurement in education, American council on education.

- Carlson, R. D. (1993). Computer Adaptive Testing: A Shift in the Evaluation Paradigm. *Journal of Educational Technology Systems*, 22(3), 213-224.
- Davey, T., & Nering, M. L. (2002). Controlling item exposure and maintaining item security. In W. C. Ward (éd.), *Computer-based testing, building foundation for future assessments*. New Jersey: Lawrence Erlbaum Associates.
- Dionne, E. (1999). *État de l'application du constructivisme en sciences physiques 416-436 dans une commission scolaire québécoise*. Mémoire de maîtrise inédit, Université de Montréal.
- Drasgow, F., & Mattern, K. (2006). New tests and new items: opportunities and issues. In R. K. Hambleton (éd.), *Computer-based testing and the internet: issues and advances* (pp. 59-89). West Sussex, England: John Wiley & Sons.
- Drasgow, F., & Olson-Buchanan, J. B. (1999a). Blood, sweat, and tears: Some final comments on computerized assessment. In J. B. Olson-Buchanan (éd.), *Innovations in computerized assessment* (pp. 249-254). Mahwah, NJ: Lawrence Erlbaum Associates.
- Drasgow, F., & Olson-Buchanan, J. B. (éds). (1999b). *Innovations in Computerized Assessment*. New Jersey: Lawrence Erlbaum Associates.
- English, R. A., Reckase, M. D., & Patience, W. M. (1977). Application of tailored testing to achievement measurement. *Behavior research methods and instrumentation*, 23(3), 341-347.
- Epstein, M. L., Lazarus, A. D., Calvano, T. B., Matthews, K. A., Hendel, R. A., Epstein, B. B., et al. (2002). Immediate feedback assessment technique promotes learning and corrects inaccurate first responses. *The Psychological Record*, 52, 187-201.
- Glowacki, M. L., McFadden, A. C., & Price, B. J. (1995). *Developing Computerized Tests for Classroom Teachers: A Pilot Study*. Paper presented at the Annual Meeting of the Mid-South Educational Research Association, Biloxi, MS, November 8-10, 1995.
- Harris, M. M. (2006). Internet testing: the examinee perspective. In R. K. Hambleton (éd.), *Computer-based testing and the internet: issues and advances* (pp. 115 et suiv.). West Sussex, England: John Wiley and Sons.
- Kingsbury, G. G., & Houser, R. L. (1999). Developing computerized adaptive tests for school children. In J. B. Olson-Buchanan (éd.), *Innovations on computerized assessment*. New Jersey: Lawrence Erlbaum Associates.
- Laurier, M. (1999). Testing adaptatif et évaluation des processus cognitifs. In B. Noël (éd.), *L'évaluation des compétences et des processus cognitifs: modèles, pratiques et contextes* (pp. 255-266). Bruxelles: DeBoeck Université.
- Luecht, R. M. (2006). Operational issues in computer-based testing. In R. K. Hambleton (éd.), *Computer-based testing and the internet: issues and advances* (pp. 91-114). West Sussex, England: John Wiley & Sons.
- McBride, J. R., & Martin, J. T. (1983). Reliability and validity of adaptive ability tests in a military setting. In A. Press (éd.), *New horizons in testing* (pp. 223-226). New York.
- McMillan, J. H. (2001). *Classroom assessment: Principles and practice for effective instruction* (2^e éd.). Needham Heights, MA: Allyn and Bacon.
- Mills, C. N., Potenza, Maria T., Fremer, John J., Ward, William C. (2002). *Computer-based testing building the foundations for future assessments*. New Jersey: Lawrence Erlbaum Associates.
- Ministère de l'Éducation du Québec (2003). *Programme de formation de l'école québécoise, enseignement secondaire, premier cycle*. Québec.

- Natal, D. (1998). *On-Line Assessment: What, Why, How*. Paper presented at the Technology Education Conference, Santa Clara, CA, May 6, 1998. Des exemples de tests contenus dans cette communication se trouvent dans le site: [<http://www.imagenmm.com/testing/>].
- Neuman, G., & Baydoun, R. (1998). Computerization of Paper-and-Pencil Tests: When Are They Equivalent? *Applied Psychological Measurement*, 22(1), 71-83.
- Nonnon, P. (2005). L'ExAO: bilan, développements récents et perspectives. *Skholé, hors-série*(2), 15-20.
- Olsen, J. B., et al. (1989). Comparisons of Paper-Administered, Computer-Administered and Computerized Adaptive Achievement Tests. *Journal of Educational Computing Research*, 5(3), 311-326.
- Powell, Z.-H. E. (1994). The Psychological Impacts of Computerized Adaptive Testing Methods. *Educational Technology*, 34(8), 41-47.
- Rabinowitz, S., & Brandt, T. (2001). *Computer-Based Assessment: Can It Deliver on Its Promise? Knowledge Brief*. San Francisco, CA: WestEd. (INFORMATION ANALYSES – State-of-the-Art Papers, Research Summaries, Reviews of the Literature on a Topic – ED-01-CO-0012)
- Rosenquist, A., Shavelson, R. J., & Ruiz-Primo, M. A. (2000). *On the «Exchangeability» of Hands-On and Computer-Simulated Science Performance Assessments*. CSE Technical Report. Standards, and Student Testing, Los Angeles, CA: Los Angeles Center for the Study of Evaluation, Center for Research on Evaluation, University of California. (Research/Technical R305B60002)
- Russo, A. (2002). Mixing Technology and Testing. *School Administrator*, 4(59), 6-12.
- Sands, W. A., Waters, B. K., & McBride, J. R. (1997). *Computerized adaptive testing: from inquiry to operation*. Washington: American psychological association.
- Scallon, G. (2004). *L'évaluation des apprentissages dans une approche par compétences*. Montréal: Éditions du nouveau pédagogique.
- Schnipke, D. L., & Scrams, D. J. (1999). *Item Theft in a Continuous-Testing Environment: What Is the Extent of the Danger? Law School Admission Council Computerized Testing Report*. Princeton, NJ: Law School Admission Council. (Research/Technical Series)
- Shermis, M. D., & Burnstein, J. C. (2003). *Automated essay scoring, a cross-disciplinary perspective*. New Jersey: Lawrence Erlbaum Associates.
- Stephens, D. (2001). Use of Computer Assisted Assessment: Benefits to Students and Staff. *Education for Information*, 19(4), 265-275.
- Sympson, J. B., & Hetter, R. D. (1985). *Controlling item exposure conditional on ability in computerized adaptive testing*. Paper presented at the Proceedings of the 27th annual meeting of the Military Testing Association, San Diego.
- Trentin, G. (1997). Computerized Adaptive Tests and Formative Assessment. *Journal of Educational Multimedia & Hypermedia*, 6(2), 201-220.
- Wainer, H., Dorans, N. J., Eignor, D., Flaugher, R., Green, B. F., Mislevy, F. J., Steinberg, L., & Thissen, D. (2000). *Computerized adaptive testing: A primer* (2^e éd.). New Jersey: Lawrence Erlbaum Associates.
- Wang, A. C.-h., & Chuang, C.-l. (2002). Applicable Adaptive Testing Models for School Teachers. *Educational Media International*, 39(1), 55-59.