

# MISE À L'ESSAI D'UN MODÈLE ÉDUMÉTRIQUE D'ÉVALUATION DES APPRENTISSAGES SCOLAIRES

MICHELINE BERCIER-LARIVIÈRE

*Conseil scolaire de district catholique du Centre-Est de l'Ontario*

RENÉE FORGETTE-GIROUX *Université d'Ottawa*

**RÉSUMÉ.** Le présent article propose un modèle éduométrique de qualité des résultats d'évaluation des apprentissages en salle de classe. De l'avis des participants, les quatre composantes du concept de justesse sont toutes pertinentes à l'obtention de résultats justes. Bien que modeste, l'expérience confirme la compatibilité du modèle avec la nature contemporaine de l'apprentissage. De ce point de vue, il s'avère supérieur aux critères de validité et de fidélité présentement utilisés en éducation.

**ABSTRACT.** This article presents an edumetric Classroom Assessment Result Quality Model. In the opinion of participants, the four elements of the concept of "soundness" are all equally important in obtaining sound results. Although modest in scope, the experiment confirms that the Model is compatible with modern learning. In this respect, the concept of soundness can be shown to be superior to the criteria of validity and reliability which are presently used in education.

## INTRODUCTION

Un modèle théorique est une représentation simplifiée d'une réalité complexe; il sert à communiquer une vision spécifique, organisée selon une logique particulière. En éducation, c'est le modèle de la psychométrie qui a toujours été utilisé en évaluation, même s'il ne convient pas à la nature des apprentissages et s'il est inopérable dans le contexte de la salle de classe. En effet, la validité<sup>1</sup> et la fidélité<sup>2</sup> sont des critères de qualité des résultats d'une mesure d'éléments unidimensionnels, stables et distribués de manière normale chez les sujets. Ces deux indices sont tout à fait appropriés à la mesure de traits ou de construits psychologiques, mais les apprentissages, eux, sont complexes, en constante évolution et l'enseignement vise leur plein développement. Il n'est donc pas toujours pertinent de recourir à la validité et à la fidélité pour établir la qualité de l'évaluation des apprentissages. L'éducation devrait plutôt se référer à un modèle adapté à la nature des apprentissages et à sa fonction pédagogique, c'est-à-dire un modèle éduométrique.<sup>3</sup>

Bien que cette réalité ait été discutée depuis le milieu des années 60, les enseignants disposent de peu de structure d'appui pour leurs démarches d'évaluation des apprentissages en salle de classe afin d'en assurer la qualité des résultats. Les lacunes rapportées dans la documentation au sujet des pratiques d'évaluation des enseignants aident à mieux comprendre cette problématique. Après avoir résumé brièvement ces informations, la présente recherche tente de remédier à cette situation en proposant un modèle éducatif actuellement en développement.

## PRATIQUES D'ÉVALUATION DES ENSEIGNANTS

Ce n'est que récemment que l'on a commencé à s'interroger sur l'utilisation de l'évaluation en salle de classe et sur sa contribution à une activité pédagogique saine et efficace. Les valeurs et les croyances personnelles établies depuis longtemps et ancrées dans les pratiques de chacun continuent de déterminer leur manière de faire (Pryor, 1992; Shepard, Flexer, Hiebert, Marion, Mayfield & Weston, 1996). Malgré la bonne intention des enseignants, il existe encore des pratiques d'évaluation qui rendent les élèves mal à l'aise et leur font honte (Bercier-Larivière, 1994). Comme le rapporte Pilcher, « les parents, les enseignants et la plupart des élèves accordent beaucoup d'importance aux notes; c'est pourquoi on les utilise pour les contraindre à travailler en classe » (1994, p. 87). Perrenoud utilise presque les mêmes termes : « derrière les notes, parents et enseignants évoquent la réussite ou l'échec scolaire, constamment à l'horizon » (1998, p. 79). Les élèves jouent un rôle négligeable dans l'évaluation, si ce n'est d'accomplir les tâches qui leur sont assignées (Ferrara, Goldberg & Mctighe, 1995; Gardner, 1992; Simon, Forgette-Giroux et Bercier-Larivière, 1993). La rétroaction qu'ils en reçoivent se limite souvent à une simple note (Bangert-Drowns, Kulik et Kulik, 1991; Baron, 1991; Dempster, 1992; Dorr-Bremme et Herman, 1986; Dumoulin, 1991; Stiggins et Conklin, 1992; Wilson, 1990).

Même si les enseignants disent évaluer les niveaux de pensée supérieurs tels que les capacités d'analyse et de synthèse des élèves, l'examen des instruments d'évaluation qu'ils utilisent révèle la présence presque exclusive d'items fondés sur le rappel de connaissances (De Pascale, 1990; Fleming et Chambers, 1983). L'évaluation formative, au vrai sens du terme, est peu pratiquée; les enseignants disent attribuer des notes à toutes formes d'évaluation (Bercier-Larivière, 1998; Perrenoud, 1998). Ces informations, consignées sous forme de notes, servent rarement à réorienter l'enseignement; elles sont gardées en registre en vue de préparer les bulletins (Forgette-Giroux, Simon et Bercier-Larivière, 1996; Herman et Klein, 1996).

Les divers aspects de l'évaluation relevés dans la documentation nous renseignent indirectement sur la capacité des enseignants à déterminer correctement les habiletés à évaluer, à construire des tâches d'évaluation

congruents à ces habiletés et à les noter. Actuellement, l'évaluation constitue souvent une activité détachée de l'enseignement; les résultats ne sont pas réinvestis dans l'activité pédagogique et les enseignants se servent de l'évaluation, entre autres, pour influencer sur les agissements de leurs élèves. Sans modèle de qualité des résultats d'évaluation des apprentissages en salle de classe, il n'y a pas de base pour inculquer aux enseignants les principes fondamentaux d'interventions évaluatives, les points critiques sur lesquels ils doivent se concentrer et les failles à éviter. Sans fondement philosophique, l'évaluation des apprentissages en salle de classe demeure loin du motif premier qui devrait motiver d'entreprendre une démarche en ce sens : situer l'élève dans son développement, reconnaître ses forces et ses faiblesses et décider du meilleur moyen de poursuivre son cheminement vers l'acquisition des compétences visées.

Le concept de justesse des résultats offre une solution possible à cette absence de modèle de qualité des résultats d'évaluation des apprentissages en salle de classe. Il propose une conception de l'évaluation congruente avec la nature moderne des apprentissages scolaires, en vue de soutenir des activités d'évaluation susceptibles de renseigner sur la manière d'apprendre de l'élève, de fournir des informations précises sur ses forces et ses faiblesses et de situer son développement par rapport à des normes établies. Qu'il s'agisse d'évaluation formative, sommative ou certificative, un résultat juste est fondé, sensé, représentatif, signifiant et vrai, entaché du moins d'erreur possible. C'est un résultat auquel il est possible de faire confiance afin de situer un élève dans son développement. Un résultat d'évaluation juste contribue à assurer le succès de l'élève et mène à des jugements éclairés et à des décisions adéquates quant au cheminement préférable de chaque élève.

#### LE CONCEPT DE JUSTESSE DES RÉSULTATS

L'opérationnalisation de cette conception de l'évaluation fait ressortir les aspects critiques d'une évaluation de qualité. La démarche engagée, respectueuse des quatre composantes établies dans le cadre conceptuel (Bercier-Larivière, 1998; Bercier-Larivière et Forgette-Giroux, 2000), s'articule autour des éléments présentés dans la figure 1:

- 1) l'agencement optimal des habiletés visées et du type de tâche choisie,
- 2) le lien avec l'activité pédagogique dans laquelle s'insère l'évaluation,
- 3) la transparence de l'activité évaluative pour l'élève,
- 4) l'absence de désavantage circonstanciel.

La justesse des résultats est un concept de nature qualitatif. Selon le respect des composantes qui le constituent, un résultat d'évaluation est plus ou

moins juste, donc plus ou moins fiable pour situer un élève dans son développement. La référence à cette conception d'une évaluation de qualité permet à l'enseignant de réunir des conditions favorables à un résultat de qualité, mais aussi de nuancer sa confiance dans les résultats obtenus, après coup. Des lacunes au niveau de la première composante mettraient en péril la justesse des résultats, peu importe la rigueur appliquée aux trois autres composantes. Toutefois, si l'une ou l'autre des composantes est moins bien respectée que les autres, quelle qu'elle soit, l'image finale de l'apprentissage évalué en sera moins juste.

| POUR UN RÉSULTAT D'ÉVALUATION JUSTE   |   |
|---|---|
| <p>1. PERTINENCE DE LA TÂCHE D'ÉVALUATION PAR RAPPORT AUX HABILITÉS VISÉES</p> <ul style="list-style-type: none"> <li>• Tâche et critère(s) d'évaluation congruents avec les diverses habiletés visées</li> <li>• Tâche signifiante, adaptée, claire, techniquement correcte et nouvelle</li> </ul>   |   |
| <p>2. COHÉRENCE AVEC L'ACTIVITÉ PÉDAGOGIQUE</p> <ul style="list-style-type: none"> <li>• Activités d'apprentissage préalables avec mise en valeur des mêmes critères d'évaluation</li> <li>• Tâche familière, mise en situation appropriée et aide disponible</li> <li>• Schème d'interprétation approprié au type d'évaluation et permettant le réinvestissement des résultats</li> <li>• Possibilité de reprise</li> </ul>      | <p>3. TRANSPARENCE DE L'ACTIVITÉ ÉVALUATIVE</p> <ul style="list-style-type: none"> <li>• Participation de l'élève à la démarche d'évaluation</li> <li>• Précisions sur l'objet et les modalités de l'évaluation</li> <li>• Schème d'évaluation familier aux élèves</li> <li>• Communication des résultats en termes de forces et de faiblesses</li> <li>• Retour sur l'activité d'évaluation en classe</li> </ul> |
| <p>4. ABSENCE DE DÉSAVANTAGE CIRCONSTANCIEL</p> <ul style="list-style-type: none"> <li>• Conditions propices à la concentration</li> <li>• Procédure équitable et habituelle</li> <li>• Conscience de facteurs néfastes à l'apprentissage (fatigue, préférences personnelles, rendement antérieur, etc.)</li> <li>• Collaboration contrôlée</li> <li>• Temps amplement suffisant</li> <li>• Présence pendant préalable</li> </ul> |   |

FIGURE 1. Composantes du concept de justesse des résultats (Bercier-Larivière, 1998, p. 39).

Afin de vérifier la viabilité du concept tel que défini dans la figure 1, il est crucial de le confronter à la réalité de la salle de classe. Le présent article porte donc sur la mise à l'essai de ce modèle et tente de répondre à la question suivante:

La pertinence de la tâche d'évaluation par rapport aux habiletés visées, la cohérence de celle-ci avec l'activité pédagogique, la transparence de l'activité évaluative et l'absence de désavantage circonstanciel rendent-elles compte de l'ensemble des éléments pouvant intervenir sur la justesse des résultats d'évaluation des apprentissages en salle de classe?

L'article expose d'abord le contexte de l'expérimentation et l'accumulation de données pertinentes, puis analyse et interprète les informations recueillies. Finalement, il discute de la pertinence du concept de justesse des résultats

et souligne quelques pratiques contraires à cette conception de l'évaluation. Il propose également quelques pistes de recherche susceptibles de faire évoluer le concept de justesse des résultats et de mieux comprendre sa portée et sa transdisciplinarité.

## **CONTEXTE DE MISE À L'ESSAI**

Compte tenu de la question de recherche, l'approche qualitative a été privilégiée parce qu'elle est plus sensible aux circonstances, à la nature des informations recueillies et à l'expérience des participants. Le recrutement des participants et la démarche adoptée pour collecter les données sont donc typiques de cette approche.

### ***Recrutement des participants***

Le concept de justesse des résultats aurait pu être expérimenté à n'importe quel niveau scolaire; ce sont toutefois les enseignants de la première à la neuvième année inclusivement, oeuvrant dans deux écoles ontariennes de langue française, qui constituèrent la population cible. Les directions d'école ont été rejointes par l'entremise d'un surintendant du conseil scolaire. Là où des enseignants étaient intéressés par ce projet, les directions d'école ont fixé un moment propice pour la première rencontre. À cette occasion, les candidats présents ont pris connaissance du projet de recherche et des exigences de la participation. En échange de leur expertise ils bénéficiaient d'un contexte favorable à l'épanouissement de leurs compétences à évaluer les apprentissages en salle de classe.

Cinq volontaires ont ainsi été recrutés. Ces personnes provenaient de deux différentes écoles d'un conseil scolaire ontarien de langue française, à une heure de distance l'une de l'autre; Christine et Kevin (noms fictifs) de la première école, Marie, Matilde et Marie-Chantal (noms fictifs) de la deuxième école. Le profil professionnel de ces cinq participants convainc du caractère plutôt ordinaire des participants; leur formation relativement récente, c'est-à-dire durant les années 90, reflète de manière générale la compétence des enseignants en exercice. Exception faite de Kevin, à ses premières armes dans la profession, elles ont toutes quatre ou cinq années d'expérience en salle de classe.

### ***Collecte des données***

L'approche naturaliste telle que décrite par Lincoln et Guba (1985) a permis de contrer l'effet de désirabilité qui caractérise souvent les méthodes indirectes de collecte des données (Gravel, 1991; Higgins et Rice, 1991; Stiggins, 1990; Van Nieuwenhoven et Jonnaert, 1992). La souplesse de cette approche a favorisé l'ajustement au contexte (Poupart, Deslauriers, Groulx, Laperrière, Mayer, Pires, 1997) et le respect de la complexité de l'activité pédagogique

étudiée (Bercier-Larivière, 1994; Bercier-Larivière et Forgette-Giroux, 1995; Rogers, 1991; Tobin, Butler-Kahle, et Fraser, 1990).

La confiance mutuelle entre les parties impliquées s'avérait cruciale pour la crédibilité des données recueillies (Stake, 1994). Les candidats ne devaient pas imaginer qu'il s'agissait d'une évaluation de leurs compétences, ni d'un parasitage de leurs activités, mais bien d'un partenariat en vue d'une compréhension approfondie du phénomène étudié (Lincoln et Guba, 1985; Pourtois et Desmet, 1988). Cette confiance s'est développée à travers des rencontres informelles, individuelles ou en équipe, selon la disponibilité, le désir et les besoins de chacun. Au début, les échanges ont porté sur des notions fondamentales d'évaluation des apprentissages, afin de s'entendre sur la terminologie à utiliser, puis se sont étendues aux activités d'évaluation régulières en salle de classe. Tout au cours de la démarche, les méprises quant au sens des termes utilisés ont vite été corrigés. Les préambules à l'opérationnalisation du concept de justesse des résultats ont duré près de 20 heures, étalées sur environ deux mois.

La vulgarisation des composantes du concept de justesse des résultats s'est faite par des ateliers de travail. Les participants ont assisté d'abord à de brèves présentations exposant les composantes du concept, puis ils ont tenté ensuite, avec l'aide de la chercheuse, d'en imaginer l'application à des situations fictives d'évaluation. Les discussions entourant chaque exemple d'application ont mis en valeur les composantes du concept de justesse des résultats et leur harmonisation; elles avaient pour but d'amener graduellement les participants à en voir la mise en pratique. L'activité s'est étendue ensuite à quelques situations fictives ou réelles d'évaluation pertinentes à l'activité pédagogique quotidienne de chacun. Le but ultime était de les rendre à l'aise avec cette conception de qualité des résultats pour que leur opinion soit crédible à la fin du projet. La figure 2 fait état du nombre de rencontres consacrées à l'initiation et à l'expérimentation du concept de justesse des résultats, de leur durée respective et de leur objet.

Cette mise à l'essai a été documentée au jour le jour au moyen d'observations notées dans un journal de bord. Les indices extérieurs comme les réactions, les commentaires, les discussions et les échanges lors des ateliers de travail et des observations dans la salle de classe y ont été inscrits, de même que la réaction des élèves. Le déroulement des activités, les difficultés rencontrées et les solutions adoptées, l'atmosphère de travail et les péripéties pertinentes au projet y ont aussi été recueillies (Bogdan et Biklen, 1992; Miles et Huberman, 1994).

Toutes les informations se rapportant à la mise en application des composantes du concept de justesse ont été systématiquement colligées dans un cahier technique, accompagnées de pièces témoins, appuyées parfois de photographies prises en salle de classe et de copies d'élèves. Après application,

## Mise à l'essai d'un modèle éduométrique d'évaluation

chacune de ces évaluations a été analysée en détail pour faire ressortir la manière dont chaque composante du concept de justesse avait pu être respectée. Ce traitement préliminaire a servi de base aux entretiens individuels menés à la fin du projet. Ceux-ci portaient sur la pertinence et l'utilité des composantes, les aspects les plus difficiles à appliquer selon l'expérience de chacun et les répercussions tant positives que négatives de cet essai sur l'activité pédagogique en général. Ces protocoles ont été transcrits intégralement, suivis de la préparation de synopsis et de leur soumission aux participants.

| ECOLE 1                          |                                  | ECOLE 2                         |                                 |                               |
|----------------------------------|----------------------------------|---------------------------------|---------------------------------|-------------------------------|
| CHRISTINE                        | KEVIN                            | MARIE                           | MATILDE                         | MARIE-CHANTAL                 |
| rencontre en équipe, 70 min.     |                                  | rencontre en équipe, 1,5 heures |                                 |                               |
| rencontre en équipe, 70 min.     |                                  | rencontre indiv. 1 heure        | rencontre indiv. 1 heure        | rencontre indiv. 1,5 heures   |
| observation en classe 3 heures   | observation en classe 2,5 heures | rencontre indiv. 90 min.        | rencontre indiv. 45 min.        | rencontre indiv. 40 min.      |
| observation en classe 3,5 heures | 90 min.                          | rencontre indiv. 30 min.        | rencontre indiv. 30 min.        | rencontre indiv.              |
| observation en classe 1,5 heures | 3,5 heures                       | observation en classe 30 min.   | rencontre indiv. 1 heure        | rencontre indiv.              |
| rencontre en équipe, 75 min.     |                                  | rencontre indiv. 20 min.        | rencontre en équipe, 1,5 heures |                               |
| entrevue 2 30 min.               | entrevue 2 30 min.               | rencontre indiv. 1 heure        |                                 | observation en classe 40 min. |
|                                  |                                  | entrevue 2 105 min.             | entrevue 2 45 min.              | entrevue 2 50 min.            |

FIGURE 2. Répartition et durée respective des diverses activités de mise à l'essai des composantes du concept de justesse des résultats.

### Situations d'évaluation

Le besoin d'outils d'évaluation pour les activités régulières d'évaluation a servi de matière première à cette mise à l'essai (Erickson, 1986); les participants ont choisi une activité pédagogique réelle à laquelle ils voulaient appliquer le concept de justesse des résultats dans leur classe. L'opérationnalisation des composantes du concept, ou leur confrontation à

diverses situations réelles, selon les besoins et l'expérience de la salle de classe, a permis d'en étudier la pertinence, la cohérence et la clarté et de les réajuster par la suite. Ces cinq situations sont les suivantes:

**SITUATION A: DÉMARCHE DE RECHERCHE SCIENTIFIQUE, 5<sup>E</sup> ANNÉE, CLASSE DE CHRISTINE**

- Évaluation formative durant l'élaboration en dyade d'un projet de recherche scientifique, de l'expérimentation et de la préparation d'un présentoir.
- Évaluation par les pairs à la suite d'une présentation orale (présentoir, expression orale, contenu) à l'aide de cartons-notes de couleurs.
- Évaluation mutuelle de la collaboration des coéquipiers, selon quatre niveaux de performance.
- Épreuve papier-crayon.

**SITUATION B: COMPRÉHENSION D'UN TEXTE ÉCRIT, 6<sup>E</sup> ANNÉE, CLASSE DE KEVIN**

- Épreuve papier-crayon portant sur divers extraits au sujet de la sécurité à bicyclette, les ceintures de sécurité en automobile et l'utilisation de l'électricité.

**SITUATION C: CONCEPT DE MULTIPLICATION, 3<sup>E</sup> ANNÉE, CLASSE DE MARIE**

- Évaluation formative et documentation de l'aide nécessaire durant un exercice en quatre volets : maîtrise du concept de la multiplication, maîtrise des diverses stratégies de multiplication, résolution d'une multiplication en situation de problème et communication de la réponse.
- À partir d'une feuille-bilan préparée à cette fin, inventaire par l'élève de ses forces quant aux 4 volets de l'exercice, suivi d'exercices de consolidation pour les volets faibles.
- Épreuve papier-crayon.

**SITUATION D: APPLICATION DE PRINCIPES DE SCIENCES ET DE MATHÉMATIQUES À LA SOLUTION D'UN DÉFI TECHNOLOGIQUE, 8<sup>E</sup> ANNÉE, CLASSE DE MATILDE**

- Évaluation par les pairs à la suite d'une présentation orale (prototype, expression orale, contenu) à l'aide de cartons-notes de couleurs.
- Évaluation mutuelle de la collaboration des coéquipiers selon une grille en quatre niveaux (participation, contribution, manipulation, rangement, utilisation du temps, persévérance).
- Appréciation globale du rapport individuel écrit sur le projet.

**SITUATION E: CONSTRUCTION D'UN MODÈLE RÉDUIT D'HÉLICOPTÈRE AVEC MOTEUR, 5<sup>E</sup> ANNÉE, CLASSE DE MARIE-CHANTAL.**

- Autoévaluation des habitudes de travail à partir d'une grille descriptive à quatre niveaux de performance (exploration, collaboration, manipulation, rangement, utilisation du temps, persévérance).

- Évaluation du prototype présenté (structure, profil, fonctionnement) .
- Appréciation globale du plan dessiné.

Dans le cas présent, cinq situations différentes d'évaluation ont été estimées suffisantes pour une première mise à l'essai. Cette limite établie *a priori* se voulait une manière de respecter les participants et de ne pas abuser de leur disponibilité, ni dénaturer leur contexte habituel de travail. De plus, ces cinq situations apparaissaient fort valables pour inspirer les meilleures stratégies à adopter en vue de faire évoluer le concept de justesse des résultats.

### RÉSULTATS DE LA MISE À L'ESSAI

Chez les participants, les composantes du concept de justesse des résultats ont été perçues comme un idéal, une référence pour réduire autant que possible les influences nuisibles sur la justesse des résultats. Ils reconnaissent que les quatre composantes comportent des éléments pouvant influencer sur le degré de confiance à accorder à un résultat pour situer correctement un élève dans son développement. Ils sont unanimes pour dire que le concept de justesse pourrait être utile à n'importe quel enseignant, novice ou chevronné, et à n'importe quel niveau scolaire.

Malgré cette position, ils ne croient pas qu'il soit possible d'appliquer toutes les composantes en tout temps, et ce, pour plusieurs raisons. Il y a d'abord l'impossibilité de réunir toutes les conditions nécessaires à l'absence de désavantage circonstanciel pour tous les élèves au même moment. Même s'ils essaient de créer une atmosphère propice à la concentration, ils peuvent difficilement considérer les conditions personnelles de chacun, parce que la plupart du temps, ils ne savent pas ce que vivent les élèves en dehors de l'école, sauf, après coup, si la situation dégénère en crise. De plus, il serait impossible de coordonner les activités d'évaluation à un moment opportun pour chacun sans apporter d'autres sources d'erreurs.

Les propos des participants lors de l'entrevue finale démontrent que le temps leur paraît être un obstacle à l'application du concept de justesse des résultats. « On a des horaires corsés par des cloches, si on pouvait enlever des cloches pour faire de l'évaluation, des fois ça ferait du bien » (Christine, entrevue finale, ligne 103). « L'activité de robotique qu'on a fait, c'était une activité d'apprentissage. . . moi j'ai juste faite [sic] une activité de robotique puis c'était mon activité d'évaluation. Si j'en avais faite [sic] d'autres, je serais pas arrivée (Matilde, entrevue finale, ligne 81). « C'est le manque de matériel, c'est le temps, le temps aussi pour les évaluer, vu que chacun travaillait individuellement, puisqu'on travaillait juste quatre [élèves] à la fois, bien je devais prendre leur temps des récréations pour les évaluer et pour eux aussi,

[il] fallait qu'ils assistent à l'évaluation, alors c'est pas tout le monde qui était bien content de manquer une récréation pour évaluer un projet » (Marie-Chantal, entrevue finale, ligne 20).

Les commentaires relatifs à l'expérimentation portent surtout sur les caractéristiques particulières du concept de justesse des résultats ou les éléments qui le distinguent des pratiques habituelles d'évaluation. Ces aspects, souvent moins bien respectés lors de la mise en application et notés dans le bilan de la situation documentée, ont été discutés lors des entrevues finales.

Au niveau des habiletés visées, de bien les identifier, je me suis aperçu que c'était une de mes faiblesses; que la tâche était bien importante mais que peut-être je [n'] identifiais pas les habiletés correctement, puis assez clairement (Chantal, ligne 9)

Je trouve que . . . c'est un petit peu difficile de bien définir la tâche et les habiletés quand il y en a plusieurs, compte tenu de l'intégration. C'est à ce moment-là qu'il faut vraiment que tu comprennes, c'est quoi exactement que je vise, ici? Donc c'est bien beau les projets d'intégration, il y a peut-être 40 habiletés mais, lesquelles? Puis c'est quelles, lesquelles que tu vas évaluer? Ça c'est difficile, alors c'est quelque chose que j'essaie de voir. Je disais bon, l'élève va développer ci, ci, ci, ça, mais on [ne] prenait peut-être pas le temps de le développer (Chantal, ligne 53)

Puis juste les indicateurs de performance, bien, je [n']en avais jamais développé (Chantal, ligne 95).

L'évaluation par les pairs, . . . quand c'est avec des couleurs c'est beaucoup mieux. C'est tout de suite. T'as pas de calculs à faire. (Matilde, ligne 153).

Quand je les fais travailler en équipe, jusqu'à quel point on peut bien évaluer le travail de chacun? (Chantal, ligne 66)

La transparence est aussi perçue comme une composante parfois difficile à appliquer. Marie, enseignante de troisième année, pense que « c'est beaucoup plus difficile chez les jeunes d'en tenir compte. C'est réaliste, je pense, parce qu'il y a des choses que tu peux faire, mais tu [ne] pourras pas atteindre tous les élèves » (entrevue finale, Marie, ligne 103).

## DISCUSSION

Les réactions et les opinions au sujet de la pertinence du concept de justesse des résultats exige de regarder les choses du point de vue des participants. Les entorses faites à leurs pratiques habituelles d'évaluation révèlent-elles des incompatibilités réelles entre le concept de justesse des résultats et l'activité pédagogique « normale » ou plutôt des contradictions entre différentes conceptions de l'évaluation? L'examen approfondi des situations

jugées problématiques vise donc à établir si elles nuisent à la justesse des résultats, si elles sont incompatibles au contexte d'enseignement/apprentissage en salle de classe ou si elles mettent simplement en évidence un malaise résultant de conceptions différentes de l'évaluation. C'est dans cette optique que les résultats ont été analysés.

Jusqu'à maintenant, les enseignants se préoccupaient peu des habiletés visées par leurs interventions. Tant pour l'enseignement et l'apprentissage que pour l'évaluation, ils se laissaient guider par un contenu à couvrir. En plaçant au premier plan les habiletés ciblées par les activités d'évaluation, le concept de justesse des résultats a fait ressortir deux situations fautives d'évaluation: les tâches d'évaluation effectuées en équipe et l'évaluation de matières enseignées de manière intégrée. Ces deux pratiques familières aux participants et très convenables pour l'apprentissage (Blumenfeld, Marx, Soloway, et Krajcik, 1996; Herman, Ashbacher et Winters, 1992; Webb, Nemer, Chizhik et Sugrue, 1998), constituent un problème d'évaluation.

Même si des auteurs (Harnish, 1994; Lussier et Turner, 1995; Mislevy, 1992, 1993; Thouin, 1993; Wiggins, 1992) affirment que l'évaluation en équipe est une suite logique à l'apprentissage en équipe, d'autres reconnaissent la situation comme étant problématique (Blumenfeld et al., 1996; Herman et al., 1992; Kulm, 1994; Leder, 1992; Webb, Nemer, Chizhik et Sugrue, 1998), parce que l'apprentissage des habiletés propres à une discipline et l'apprentissage de la collaboration sont deux gammes d'habiletés différentes. D'une part, les habiletés sociales ne sont généralement pas développées également chez tous les membres de l'équipe et peuvent masquer les autres habiletés développées en parallèle. Puisqu'il est difficile de tout évaluer à la fois, il serait préférable d'établir des priorités et décider à l'avance quel groupe d'habiletés sera évalué et de le faire sur une base individuelle. Dans tous les cas, les habiletés visées par l'évaluation doivent être précisées et les notes attribuées individuellement. D'autre part, le partage des responsabilités peut permettre à des coéquipiers d'être exemptés de démontrer leur maîtrise de certains apprentissages. Une seule note attribuée à l'équipe n'est habituellement pas méritée par tous les membres du groupe; cependant, si chacun reçoit une note personnelle, la collaboration peut sembler une perte de temps. L'évaluation par des tâches individuelles connexes à l'activité d'apprentissage effectuée en groupe force chacun à trouver profit dans la collaboration; c'est une façon de distinguer l'apprentissage de son évaluation, tout en favorisant des résultats plus justes, sans perdre les avantages d'apprendre en collaboration.

Par ailleurs, les enseignants voient les projets comme de bonnes occasions de faire de l'intégration de matières; ils rassemblent des contenus jugés compatibles pour en faire un thème intéressant. L'application du concept de justesse, parce qu'il impose d'établir les habiletés visées par l'activité d'évaluation, fait émerger de vraies préoccupations: A-t-on vraiment

développé toutes les habiletés abordées? L'apprentissage est-il suffisant pour justifier une évaluation? A-t-on assez de traces se rapportant à toutes ces habiletés pour croire que le résultat est juste? L'interaction des habiletés concernées, si elles sont nombreuses et complexes, ne vient-elle pas brouiller les traces permettant de juger de l'apprentissage réalisé?

Ce n'est pas l'apprentissage en groupe ou l'apprentissage de manière intégrée qui peut porter préjudice à l'élève mais bien l'évaluation dans ces contextes. Même si les matières intégrées et les tâches effectuées en équipe comportent toutes les deux de nombreux avantages au point de vue de l'apprentissage, elles peuvent entraîner des erreurs lors de l'évaluation. En exigeant d'abord l'établissement des habiletés visées, l'application du concept de justesse des résultats fait ressortir les dangers encourus dans ces deux situations précises. Même si une tâche regroupe plusieurs habiletés, celles-ci ne sont pas nécessairement toutes assez développées pour être évaluées. Il est donc important d'établir lesquelles le sont assez pour être évaluées et comment elles le seront, tout en limitant l'intervention de celles qui ne sont pas assez développées et ne seront donc pas évaluées.

L'opinion des enseignants-participants sur la pertinence et l'intégrabilité du concept de justesse paraît très influencée par leurs pratiques habituelles et leur connaissance de moyens de faire autrement. Avant l'application du concept, tous les participants, à l'exception de Marie, avaient l'habitude d'octroyer des « points » à toutes les tâches que faisaient les élèves; ils ne distinguaient pas les situations d'apprentissage de celles d'évaluation. De plus, ils évaluaient ce que les élèves avaient retenu des contenus enseignés plutôt que les habiletés développées. Ainsi réduite à l'attribution de notes chiffrées, l'évaluation était rarement confiée aux pairs parce que considérée trop longue et fastidieuse.

Le concept de justesse des résultats, une fois bien compris, s'oppose à de nombreuses pratiques pédagogiques actuelles, pour plusieurs raisons : l'obligation de distinguer les activités d'apprentissage de celles d'évaluation, l'interdiction d'accorder des points à tout ce que font les élèves pour les encourager à travailler et les résultats exprimés en termes de forces et de faiblesses plutôt qu'en notes chiffrées. Cependant, ces incohérences entre les pratiques actuelles et les composantes du concept de justesse révèlent plutôt des pratiques qui nuisent à la justesse des résultats et non pas des faiblesses du concept comme tel. Elles sont donc vues comme des indices de conceptions différentes de l'évaluation des apprentissages et non pas des preuves du caractère disfonctionnel du concept de justesse des résultats.

L'opérationnalisation du concept de justesse des résultats a permis de préciser ses quatre composantes, de les décrire avec plus de détails et d'y intégrer des éléments complémentaires. Les quatre tableaux qui suivent présentent donc une version renouvelée des composantes du concept de justesse suite à la mise à l'essai.

## Mise à l'essai d'un modèle éducatif d'évaluation

TABLEAU 1. Première composante du concept de justesse des résultats révisée suite à la mise à l'essai

### PERTINENCE DE LA TÂCHE D'ÉVALUATION PAR RAPPORT AUX HABILETÉS VISÉES

#### HABILETÉS

- Les habiletés complexes sont décortiquées pour mieux cibler les habiletés mises en relation.
- Le recours à des habiletés complexes en voie d'être atteintes n'est pas implicitement nécessaire.

#### CRITÈRES

- Les critères sont congruents avec les habiletés ciblées, réalistes et typiques des habiletés en milieu authentique.
- Ils sont les mêmes que ceux valorisés lors de l'apprentissage.
- Ils excluent des considérations étrangères aux habiletés visées (qualité du français, calligraphie, propreté).
- Ils sont définis et communiqués en même temps que la tâche

#### TÂCHE

- La tâche d'évaluation est nouvelle et distincte de celle d'apprentissage (pas de note attribuée pour faire les activités d'apprentissage).
- Elle est appropriée aux habiletés visées, définie clairement, succinctement, en particulier pour les jeunes élèves et techniquement bien construite. (La démonstration d'habiletés complexes est impossible avec des automatismes)
- Le niveau de difficulté et la durée respecte la capacité de concentration des élèves évalués.
- Il vaut la peine d'y investir l'énergie et le temps requis pour bien la faire (elle est motivante et signifiante).
- L'exécution laisse des traces de la performance des évalués

TABLEAU 2. Deuxième composante du concept de justesse des résultats révisée suite à la mise à l'essai

### COHÉRENCE DE L'ACTIVITÉ D'ÉVALUATION AVEC L'ACTIVITÉ PÉDAGOGIQUE

- Les activités d'apprentissage précèdent l'évaluation et mettent en valeur les mêmes critères de performance.
- L'activité d'évaluation est en soi une situation d'apprentissage.
- La situation d'évaluation et la tâche sont familières.
- La mise en situation est appropriée et stimulante.
- Du matériel concret ou des ressources sont disponibles, comme pour l'apprentissage
- De l'aide est disponible pour obtenir des précisions, comme pour l'apprentissage.
- Les résultats sont exprimés de manière à pouvoir être éventuellement réinvestis (profile d'habiletés).
- Il y a possibilité de reprise.

TABLEAU 3. Troisième composante du concept de justesse des résultats révisée suite à la mise à l'essai

### TRANSPARENCE DE L'ACTIVITÉ ÉVALUATIVE

- L'activité d'évaluation est prévue d'avance.
- L'élève est prévenu de l'objet et des modalités de l'évaluation.
- Les critères d'évaluation et les indicateurs de rendement sont clairs pour l'élève et connus avec la tâche.
- Des modèles de travaux sont utilisés pour communiquer les attentes.
- Les résultats correspondent à des forces et des faiblesses (l'élève peut juger s'ils sont fondés) et sont rattachés à une appréciation de la performance (bien, très bien, exceptionnel).
- L'élève participe à l'évaluation de ses apprentissages, selon sa maturité et son niveau d'autonomie (autoévaluation, définition des critères, évaluation par les pairs, autoévaluation à l'aveugle, etc.).
- Il y a retour en classe sur l'activité d'évaluation après la notation.
- Il est possible de réajuster la note en présence d'arguments convaincants.

TABLEAU 4. Quatrième composante du concept de justesse des résultats révisée suite à la mise à l'essai

---

**ABSENCE DE DÉSAVANTAGE CIRCONSTANCIEL**

---

- Les absences de l'élève n'ont pas nuit à l'apprentissage.
  - L'enseignant est conscient des effets possiblement néfastes de la fatigue, des préférences personnelles et du rendement antérieur de chaque élève.
- 

**TRAVAIL INDIVIDUEL**

- La tâche ne comporte pas de biais.
  - Le temps alloué est amplement suffisant.
  - La collaboration entre élèves est documentée, si elle est permise.
  - La tâche est effectuée en classe, autant que possible (pas à la maison, à la récréation ni à l'heure du dîner).
  - Les conditions sont propices à la concentration (absence de : dérangement à l'horaire, d'activité spéciale, de chaleur excessive, de bavardage durant les présentations, de préoccupations personnelles sérieuses).
- 

**TRAVAIL EN ÉQUIPE**

- Le nombre de personnes convient à l'ampleur de la tâche.
- Le partage des responsabilités ne permet pas à certains membres d'être exemptés complètement certaines habiletés.
- Les partenaires sont compatibles (niveau d'engagement, motivation).
- Le regroupement est géré si les tâches sont susceptibles de comporter des stéréotypes, (par exemple, une tâche impliquant de la mécanique ou la préparation d'aliments sera effectuée par groupes unisexes).

**CONCLUSION**

Après toutes ces heures de collaboration, les participants se disent heureux de l'expérience et auraient aimé qu'elle dure plus longtemps. Une fois familiers avec le concept de justesse des résultats, puis l'avoir appliqué à une situation réelle, ils sont unanimes: le concept de justesse, avec ses quatre composantes, témoigne bien de la réalité; il rassemble des éléments qui peuvent tous intervenir sur la capacité d'un résultat d'évaluation à situer correctement un élève dans son développement. Ils le trouvent complet et fidèle à la réalité, ils n'ont pas d'ajout ou de retrait à proposer, si ce n'est qu'il devrait être intégré à toute forme d'enseignement. L'expérience menée ici confirme la viabilité du concept et permet d'orienter la manière de perfectionner le concept correspondant. Enfin, cette confrontation à la réalité de la salle de classe a permis, comme l'affirme Keeves (1988), de donner de la valeur au modèle théorique qui sous-tend ce concept.

L'exercice a permis de préciser les composantes du concept de justesse. L'énoncé des habiletés visées apparaît être la force majeure du concept de justesse des résultats en ce qu'il incite à distinguer des habiletés telles que le rappel de connaissances et l'application d'algorithmes d'autres habiletés plus complexes. Cette première composante du concept de justesse a d'abord fait ressortir deux situations d'apprentissage pouvant menacer la justesse des résultats lorsqu'elles sont utilisées aussi pour l'évaluation. Le travail en équipe fait appel à des habiletés sociales nécessaires au bon fonctionnement du groupe et l'intégration de matières risque de léser les élèves en rendant nécessaires des habiletés connexes en voie d'être atteintes.

En ce qui a trait à des recherches futures, il serait nécessaire d'approfondir les moyens pour assurer la transparence de l'évaluation pour les jeunes élèves, particulièrement si ces jeunes ne maîtrisent pas encore la lecture. Il semble que la responsabilisation des élèves dès la première année soit souhaitable, mais l'ajuster à leur degré de maturité et d'autonomie pose des défis différents.

Dans le contexte d'examens provinciaux récemment implantés en Ontario, il serait intéressant d'appliquer les composantes du concept de justesse aux résultats obtenus dans divers groupes-classes. En examinant les habiletés ciblées par les activités d'évaluation imposées, leur cohérence avec les activités pédagogiques vécues auparavant, leur transparence à l'égard de l'élève et la présence de désavantages circonstanciels, il serait possible de doser la confiance à accorder aux résultats de ces examens pour situer correctement les élèves dans leur développement. Le résultat de ces analyses pourrait tempérer les comparaisons possibles d'un groupe à l'autre et expliquer certains écarts de performance.

Cette première opérationnalisation du concept de justesse des résultats comporte une limite importante : il n'a été appliqué qu'à cinq différentes situations d'évaluation. Cependant, malgré l'écart qu'il présente dans la manière de concevoir l'activité évaluative, le concept est jugé utile, réaliste et intégrable aux activités pédagogiques de la salle de classe. Cette brève mise à l'essai a fait ressortir pour la première fois l'aspect problématique de situations d'évaluation et identifié deux freins importants à l'instauration de pratiques évaluatives plus justes: l'attribution de points à tout ce que font les élèves et l'intervention pédagogique axée sur des contenus à couvrir plutôt que sur des habiletés à développer.

L'évaluation en salle de classe est la plus déterminante pour l'expérience scolaire. Cependant, déloger des pratiques ancrées depuis longtemps ne peut se faire du jour au lendemain, surtout si les principaux intéressés n'ont ni les fondements théoriques ni le soutien nécessaires pour bien comprendre pourquoi et comment ils devraient le faire. Entre temps, il apparaît primordial de poursuivre la recherche sur le concept de justesse des résultats parce qu'il pourrait, éventuellement, devenir un schème de référence éducatif convenable. Il pourrait servir à inculquer aux enseignants les principes fondamentaux d'évaluation des apprentissages, faciliter la recherche en ce domaine et uniformiser davantage les pratiques.

#### NOTES

1. Degré selon lequel un test mesure bien ce qu'il prétend mesurer.
2. Constance avec laquelle une personne obtiendrait le même résultat de mesure dans des conditions semblables.
3. Modèle conforme à la mesure d'éléments multidimensionnels et évolutifs, visant à mettre en valeur les changements chez un même individu, selon un idéal établi, dans le but de réguler les apprentissages et de reconnaître les acquis.

## RÉFÉRENCES

- Bangert-Drowns, R., Kulik, J. et Kulik, C. (1991). Effects of frequent classroom testing. *Journal of Educational Research*, 85(2), 89-99.
- Baron, J. (1991). Strategies for the development of effective performance exercises. *Applied Measurement in Education*, 4, 305-318.
- Bercier-Larivière, M. (1994). *La motivation des élèves de 6<sup>e</sup>, 7<sup>e</sup> et 8<sup>e</sup> année et leur perception des pratiques d'évaluation des apprentissages*. Thèse de maîtrise non publiée, Université d'Ottawa, Ottawa.
- Bercier-Larivière, M. (1998). *La justesse des résultats, critère de qualité de l'évaluation des apprentissages en salle de classe*. Thèse de doctorat non publiée, Université d'Ottawa, Ottawa.
- Bercier-Larivière, M. et Forgette-Giroux, R. (1995). L'évaluation des apprentissages telle que perçue par des élèves de 6<sup>e</sup>, 7<sup>e</sup> et 8<sup>e</sup> année. *Mesure et évaluation en éducation*, 18(1), 37-58.
- Bercier-Larivière, M. et Forgette-Giroux, R. (1999). L'évaluation des apprentissages scolaires: une question de justesse. *Revue canadienne d'éducation/Canadian Journal of Education*, 24(1), 169-182.
- Blumenfeld, P., Marx, R., Soloway, E. et Krajcik, J. (1996). Learning with peers: From small group cooperation to collaborative communities. *Educational Researcher*, 25(8), 37-40.
- Bogdan, R. C. et Biklen, S. K. (1992). *Qualitative research for education. An introduction to theory and methods* (2<sup>e</sup> éd.). Toronto: Allyn and Bacon.
- Dempster, F. (1992). Using tests to promote learning: A neglected classroom resource. *Journal of Research and Development in Education*, 25, 213-217.
- DePascale, C. (1990). *Factors influencing the cognitive emphasis of classroom tests*. Doctoral Thesis, University of Minnesota, Dissertation Abstracts International, DAI-A 51/05, p. 1587.
- Dorr-Bremme, D. et Herman, J. (1986). *Assessing student achievement: A profile of classroom practices*. Los Angeles: University of California, Center for the Study of Evaluation.
- Dumoulin, J. (1991). *Etude des pratiques de l'évaluation des apprentissages*. Thèse de maîtrise, Université de Montréal.
- Erickson, F. (1986). Qualitative methods in research on teaching. Dans V. Shulman et M. Wittrock, *Handbook of research on teaching* (3<sup>e</sup> éd., pp. 119-161). New York: Macmillan.
- Ferrara, S., Goldberg, G. et McTighe, J. (1995). *Ways in which teachers communicate learning targets, criteria, and standards for performance to their students*. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco.
- Fleming, M. et Chambers, B. (1983). Teacher-made tests: Windows to the classroom. In W. E. Hathaway (Ed.), *Testing in the schools* (pp. 29-47). San Francisco: Jossey-Bass.
- Forgette-Giroux, R., Simon, M. et Bercier-Larivière, M. (1996). Les pratiques d'évaluation des apprentissages en salle de classe: Perceptions des enseignants et des enseignantes. *Revue canadienne de l'éducation/Canadian Journal of Education*, 21, 384-395.
- Gardner, H. (1992). Assessment in context: The alternative to standardized testing. In B. Gifford & M. O'Connor, *Changing assessments: Alternative views of aptitude, achievement, and instruction* (pp. 77-120). Boston: Kluwer.
- Gravel, L. (1991). *L'évaluation du rendement de l'élève dans les écoles élémentaires et secondaires de langue française de la région Nord-Est*. Document non publié, Université d'Ottawa.

- Harnisch, D. (1994). Performance assessment in review: New directions for assessing student understanding. *International Journal of Educational Research*, 21, 341-350.
- Herman, J., Aschbacher, P. et Winters, L. (1992). *A practical guide to alternative assessment*. Alexandria (VA): ASCD.
- Higgins, N. et Rice, E. (1991). Teachers' perspectives on competency-based testing. *Educational Technology Research and Development*, 39(3), 59-69.
- Keeves, J. P. (1988). Models and Model building. Dans J. P. Keeves (dir.), *Educational research, methodology, and measurement: An international handbook* (pp. 559-566). Oxford : Pergamon Press.
- Kulm, G. (1994). *Mathematics assessment. What works in the classroom*. San Francisco: Jossey-Bass.
- Leder, G. (1992). *Assessment and learning of mathematics*. Victoria, Australia: Australian Council for Educational Research.
- Lincoln, Y. S. et Guba, E. G. (1985). *Naturalistic inquiry*. Beverly Hills: Sage Publications.
- Lussier, D. et Turner, C. (1995). *Le point sur ... L'évaluation en didactique des langues*. Anjou: Centre Éducatif et Culturel.
- Miles, M. B. et Huberman, A. M. (1994). *Qualitative data analysis* (2<sup>e</sup> éd.). London: Sage.
- Mislevy, R. (1992). *Linking educational assessments. Concepts, issues, methods, and prospects*. Princeton, Educational Testing Service.
- Mislevy, R. (1993). Foundations of a new test theory. Dans N. Frederiksen, R. Mislevy et I. Bejar, (dirs) *Test theory for a new generation of tests* (pp. 19-39). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Perrenoud, P. (1998). *L'évaluation des élèves*. Paris : De Boeck.
- Pilcher, J. K. (1994). The value-driven meaning of grades. *Educational Assessment*, 2(1), 69-88.
- Poupart, J., Deslauriers, J.-P., Groulx, L.-H., Laperrière, A., Mayer, R., Pires, A. (1997). *La recherche qualitative. Enjeux épistémologiques et méthodologiques*. Montréal: Gaëtan Morin.
- Pourtois, J. P. et Desmet, H. (1988). *Épistémologie et interprétation en sciences humaines*. Liège: Pierre Mardaga.
- Pryor, E. L. (1992). *Case studies of three first-grade teachers' assessment and evaluation of literacy (grading)*. Doctoral Thesis, Kent State University. Dissertation Abstracts International, DAI-A 56/08, p. 2669.
- Rogers, T. (1991). Educational assessment in Canada : Evolution or extinction? *The Alberta Journal of Educational Research*, 37, 179-192.
- Shepard, L. , Flexer, R., Hiebert, E., Marion, S., Mayfield, V. et Weston, T. (1996). Effects of introducing classroom performance assessments on student learning. *Educational Measurement: Issues and Practice*, 15(3), 7-18.
- Simon, M., Forgette-Giroux, R. et Bercier-Larivière, M. (1993). *L'évaluation des apprentissages des élèves*. Rapport d'enquête menée auprès d'enseignants du secondaire du Conseil des Écoles Séparées Catholiques de Stormont, Dundas et Glengarry.
- Stake, R. E. (1994). Case studies. Dans N. Denzin et Y. Lincoln (dirs), *Handbook of qualitative research* (pp. 236-247). London: Sage.
- Stiggins, R. (1990). Toward a relevant classroom assessment research agenda. *The Alberta Journal of Educational Research*, 36, 92-97.
- Stiggins, R. J. et Conklin, N. F. (1992). *In teachers' hands. Investigating the practices of classroom assessment*. New York: State University of New York Press.

Thouin, M. (1993). L'évaluation des apprentissages en mathématiques: Une perspective constructiviste. *Mesure et évaluation en éducation*, 16(1-2), 47-64.

Tobin, K., Butler-Kahle, J. et Fraser, B. (dirs) (1990). *Windows into science classrooms*. New York: Palmer Press.

Van Nieuwenhoven, C. et Jonnaert, P. (1992). Approcher les représentations des enseignants du primaire à propos de l'évaluation formative. Dans D. Laveault, *Les pratiques d'évaluation en éducation* (pp. 125-138). XV<sup>e</sup> session d'étude de l'Association pour le développement de la mesure et de l'évaluation en éducation. ADMEE-Canada.

Webb, N. M., Nemer, K. M., Chizhik, A. W. et Sugrue, B. (1998). Equity issues in collaborative group assessment: Group composition and performance. *American Educational Research Journal*, 35, 607-651.

Wiggins, G. (1992). Creating tests worth taking. *Educational Leadership*, 49(8), 26-33.

Wilson, R. (1990). Classroom processes in evaluating student achievement. *The Alberta Journal of Educational Research*, 36, 4-17.

**MICHELINE BERCIER-LARIVIÈRE** a complété une maîtrise en arts et un doctorat en mesure et évaluation en éducation à l'Université d'Ottawa, après avoir enseigné au palier secondaire pendant quelques années. Elle est actuellement chef à l'imputabilité au Conseil de district catholique du Centre-Est de l'Ontario.

**RENÉE FORGETTE-GIROUX** est professeure en mesure et évaluation à la Faculté d'éducation de l'Université d'Ottawa. Ses intérêts de recherche sont l'évaluation des apprentissages: politiques, pratiques, stratégies et qualités des instruments.

**MICHELINE BERCIER-LARIVIÈRE** completed an M.A. and a Ph.D. in Measurement and Evaluation at the University of Ottawa, after teaching a number of years at the secondary level. She is currently Chief, Accountability for the Central-Eastern-Ontario French Catholic District School Board.

**RENÉE FORGETTE-GIROUX** is a professor in Measurement and Evaluation at the Faculty of Education at the University of Ottawa. Her research interests are in classroom assessment: policies, practices, strategies, and qualities of instrument.