Expérimentation d'un modèle d'évaluation certificative dans un contexte d'enseignement scientifique

Eric Dionne
Université d'Ottawa

Michel D. Laurier
Université de Montréal

Le programme québécois de science et technologie est basé sur une approche par compétences. Ce choix implique des défis importants et principalement quand l'évaluation est de nature certificative. Une des compétences à évaluer concerne l'investigation scientifique. En s'appuyant sur les travaux de Rey *et al*. (2003), nous avons conçu un modèle d'évaluation qui permet de juger du développement de cette compétence. Différentes situations d'évaluation ont été créés et administrées auprès de 560 élèves du secondaire pour vérifier si le modèle : (1) est adéquat pour mesurer le niveau de compétence des élèves et (2) se comporte de la même façon selon le contexte disciplinaire. Les résultats montrent que le modèle permet de classer les élèves selon trois niveaux de maîtrise : compétence assurée, compétence partielle et maîtrise des habiletés.

Mots clés : évaluation, compétences, investigation scientifique

The Science and Technology curriculum in the Province of Quebec, based on competencies, represents a challenge for science teachers, particularly for high-stakes assessment. Teachers who do know how to conduct hands-on assessment must deal with practical constraints. In this context, we adapted Rey's *et al.* (2003) work to construct an assessment model in relation with the Scientific Inquiry Competence. We designed different assessment situations which we administered to 560 junior high school students to verify whether a) the model is helpful in assessing learners' level of competency in scientific inquiry b) the results are comparable among disciplines on which the assessment situations are based. The results show that the model works as predicted for different learners' levels: Full competency, Partial Competency, Skill.

Key words: assessment, competencies, scientific inquiry

REVUE CANADIENNE DE L'ÉDUCATION 33, 1 (2010): 83-107
©2010 Canadian Society for the Study of Education/
Société canadienne pour l'étude de l'éducation

Depuis 2005, les enseignants[1] du secondaire ont officiellement la responsabilité de la mise en œuvre d'un nouveau programme intégré de science et technologie. Ce dernier regroupe six champs disciplinaires à savoir : physique, chimie, biologie, astronomie, géologie et technologie. De plus, ce programme est basé sur une approche dite « par compétences ». Les élèves québécois doivent, en effet, développer trois compétences disciplinaires:

(1) chercher des réponses ou des solutions à des problèmes d'ordre scientifique ou technologique;

(2) mettre à profit ses connaissances scientifiques et technologiques;

(3) communiquer à l'aide des langages utilisés en science et technologie. (Ministère de l'Éducation du Québec, 2003c, p. 268)

Afin de développer ces compétences, on suggère de placer les élèves face à des situations d'apprentissage complexes où ils devront mobiliser les ressources nécessaires afin de résoudre la situation proposée. Dans un souci de cohérence et surtout de validité, il convient donc de placer les élèves dans des situations similaires en contexte d'évaluation certificative. La question qu'on doit dès lors se poser est : comment évaluer ces compétences compte tenu de ce nouveau contexte et des multiples contraintes (accessibilité du matériel de laboratoire, disponibilité des classes-laboratoires, temps limité consacré à l'enseignement de cette matière, etc.) qui caractérisent l'enseignement de la science et de la technologie au secondaire ? Cet article relate une expérience visant à mesurer l'effet de la mise en œuvre d'un modèle inédit d'évaluation utilisé dans un contexte certificatif afin de rendre compte d'une compétence scientifique.

CONTEXTE DE L'ENSEIGNEMENT DE LA SCIENCE ET TECHNOLOGIE

Depuis quelques années, le ministère de l'Éducation, du Loisir et du Sport (MELS) s'est engagé dans une vaste opération de mise à niveau de ses programmes d'études au primaire et au secondaire. Cette opération, maintenant appelée « renouveau pédagogique » amène de grands bouleversements pour les enseignants qui doivent maintenant s'ajuster en fonction de plusieurs changements fort importants comme, entre autres,

[1] Le masculin est employé dans ce texte à titre épicène dans le seul but de l'alléger.

le passage d'une approche par objectifs à une approche par compétences, une conception de l'apprentissage basée sur le cognitivisme, le constructivisme et le socioconstructivisme, des règles de passage par cycles de deux années plutôt que par année scolaire [2] et une conception plus intégrée des contenus disciplinaires. Parmi tous les changements auxquels les enseignants doivent faire face, l'approche par compétences est certainement l'un des plus importants. Le MELS a en effet décidé que les élèves devraient maintenant développer un nombre limité de compétences sur l'ensemble du primaire et du secondaire ce qui représente au Québec onze années d'étude[3]. Par exemple, les élèves doivent développer les mêmes compétences disciplinaires en science et technologie de la première année du primaire à la dernière année du secondaire.

Le concept de compétence peut être polysémique et certains n'hésitent pas à souligner la difficulté à bien circonscrire ce concept (Boutin & Julien, 2000). En ce qui nous concerne, nous retenons la définition proposée par le MELS et qui s'applique aux ordres du primaire et du secondaire. La compétence se définit comme suit : « un savoir-agir fondé sur la mobilisation et l'utilisation efficaces d'un ensemble de ressources » (Ministère de l'Éducation du Québec, 2001, p. 4). Ainsi, la seule maîtrise des concepts[4] n'est pas suffisante dans une telle approche. En définitive, l'élève doit non seulement maîtriser les concepts, mais en faire un usage adéquat dans le bon contexte afin de résoudre une situation qui est, pour l'élève, complexe. Une telle approche convient particulièrement bien à la

---

[2] À titre d'exemple, la sanction des études se réalisait auparavant à la fin de chaque année scolaire (ex. fin de la première secondaire) alors que maintenant elle se réalise à la fin du cycle (ex. à la fin de la deuxième année pour le premier cycle du secondaire).

[3] Pour un élève qui chemine normalement, le primaire est divisée en trois cycles de deux ans ce qui représente six années d'étude. Le premier cycle du secondaire dure deux ans et ensuite l'approche est annuelle pour les trois dernières années. Le secondaire dure donc cinq ans.

[4] Nous employons l'expression « concept » au sens large ce qui inclus les notions, les techniques, les attitudes, etc. Dans le jargon ministériel québécois, nous parlerions alors de « ressources ». L'objet de cet article ne vise pas à débattre des ressemblances et des différences entre ces termes que nous considérons équivalents. Les personnes intéressées peuvent consulter l'ouvrage de Jonnaert (2002) qui aborde avec clarté ces aspects.

discipline de science et technologie. En effet, de nombreuses critiques reprochaient l'absence de mise en contexte ou, autrement dit, la distance entre ce que les élèves réalisaient en classe et la réalité quotidienne qui exige une capacité de transfert (American Association for the Advancement of Science, 1993; Fourez, 1994; Mathy, 1997). Ces critiques ne sont pas exclusives au Québec, car on peut lire les mêmes en Europe et aux États-Unis par exemple.

Ces changements au niveau pédagogique amènent forcément des changements importants au niveau de l'évaluation des apprentissages. Les enjeux sont nombreux et importants comme le souligne Dionne (2005) en les catégorisant sur quatre axes : (1) théorique, (2) méthodologique, (3) pragmatique et (4) professionnel. Dans le cadre de cette recherche, nous nous sommes intéressés au deuxième axe à savoir l'axe méthodologique afin de mieux cerner comment il était possible de juger du développement d'une compétence scientifique dans un contexte d'évaluation certificative.

Tous les changements que nous venons de décrire ont des répercussions sur le processus d'évaluation et, en particulier, au niveau des outils servant à recueillir les manifestations de ces compétences. Depuis quelques années, le MELS et les praticiens emploient des expressions comme « situation d'apprentissage » ou « situation d'évaluation » pour faire référence à certains de ces outils servant à la cueillette de l'information. Comme le souligne Beckers (2002), on doit s'assurer que les situations sollicitent la compétence visée et qu'elles soient en cohérence avec les apprentissages développées. Ainsi, il importe de fournir aux élèves des situations où ils seront en mesure de faire la démonstration de leur compétence. Or, lors d'une étude précédente (Dionne, 2000), nous avons montré que la cohérence curriculum-évaluation n'était pas toujours respectée par les enseignants de science à cause du fait, entre autres, que les situations complexes administrées aux élèves conduisaient souvent à un syndrome qualifié « de la page blanche ». Ce syndrome apparaît quand les élèves, confrontés à une situation jugée trop complexe, n'arrivent à fournir aucune indication quant à la réponse ou au processus attendu.

Dans le champ disciplinaire qui nous intéresse, science et technologie, ces changements sont qualifiés de majeurs entre autres par Potvin et Dionne (2007), Barma et Guilbert (2005) et Francoeur (2005). Par exem-

ple, les enseignants du premier cycle du secondaire avaient auparavant pour tâche d'enseigner l'écologie en première secondaire et les sciences physiques en deuxième secondaire. Depuis 2005, ils doivent enseigner un cours de science qui regroupe la chimie, la physique, la biologie, la géologie et l'astronomie et à laquelle s'ajoute la technologie. De toutes les disciplines touchées par le renouveau pédagogique, aucune autre n'a un mandat aussi lourd en termes d'intégration disciplinaire.

Le MELS impose maintenant le développement de trois compétences disciplinaires en science et technologie. La première compétence est en lien avec la démarche d'investigation scientifique et la démarche de conception en technologie. La seconde compétence aborde, quant à elle, la mise à profit des connaissances scientifiques et technologiques. Enfin, la dernière compétence est de l'ordre de la communication à l'aide des langages utilisés en science et technologie (Ministère de l'Éducation du Québec, 2003c). Les travaux que nous avons menés ont touché exclusivement aux éléments de la première compétence dans le contexte scientifique; nous y avons ajouté les éléments pertinents de la troisième compétence.

L'ÉVALUATION CERTIFICATIVE : UN CONTEXTE CONTRAIGNANT

Nous avons choisi d'expérimenter notre modèle dans un contexte d'évaluation certificative. Legendre (2005) définit l'évaluation certificative comme un « processus d'évaluation débouchant sur une décision dichotomique de réussite ou d'échec relatif à une période d'apprentissage, d'acception ou de rejet d'une promotion, de poursuite d'une action ou de l'arrêt de celle-ci » (p. 633). Une caractéristique importante associée à ce contexte est la sévérité des enjeux pour les différents acteurs et en particulier pour l'élève. Dans ce contexte, l'évaluateur doit aussi être en mesure de porter le jugement sur l'apprentissage de chaque individu, ce qui suppose de la part de l'élève un travail individuel afin d'éviter les effets de contamination. Contrairement au contexte de l'évaluation formative où l'enseignant peut et doit soutenir l'élève dans ses apprentissages, le contexte de l'évaluation certificative implique que l'enseignant offre les conditions de passation les plus uniformes possibles tout en tenant compte des besoins particuliers que pourraient

avoir certains élèves (ex. élèves qui ont un handicap physique, élèves qui possèdent des troubles du comportement, etc.). En ce sens, un effort de systématisation doit être réalisé pour s'assurer que les informations transmises aux élèves soient équivalentes pour tous.

LA NÉCESSITÉ D'EXPLORER DE NOUVEAUX MODÈLES D'ÉVALUATION

En consultant les écrits, on constate que les modèles liés à l'évaluation des situations d'apprentissage complexes sont plutôt rares. Ce constat a d'ailleurs été réalisé il y a plusieurs années déjà (Dionne, 2007; 2008; Scallon, 2004; Tardif, 2006) : « Quant au champ des productions complexes, il a été quelque peu délaissé, et la méthodologie qui s'y rattache reste le parent pauvre du domaine de l'évaluation » (Scallon, 2004, p.35). Les travaux de Brown et Shavelson (1996) de Leighton, Rogers et Maguire (1999) et de Klein *et al.* (1997) présentent des résultats intéressants concernant l'évaluation d'apprentissages scientifiques mais rarement dans un contexte de certification.

Le MELS semble d'ailleurs être non seulement au fait de ce constat mais également sensible à celui-ci puisque sa Politique d'évaluation des apprentissages (Ministère de l'Éducation du Québec, 2003b) qui dresse les grandes orientations en évaluation des apprentissages est jumelée à un plan de mise en œuvre (Ministère de l'Éducation du Québec, 2003a) qui souligne, entre autres, la nécessité de mener des recherches afin de mieux cerner la façon d'évaluer des apprentissages complexes comme des compétences.

Au cours de cette recherche, nous avons tenté de répondre à deux questions : (1) Est-ce que le modèle à trois niveaux que nous proposons permet de juger du développement d'une compétence d'investigation scientifique ? et (2) Est-ce que le modèle se comporte de la même façon selon le contexte disciplinaire ?

MÉTHODOLOGIE

Dans la présente section, nous présenterons les aspects méthodologiques sur lesquels nous nous sommes basés pour mener à bien cette recherche. Dans un premier temps, nous présenterons les outils de cueillette de données que nous avons produits à savoir les trois situations

d'évaluation (SE) en expliquant également la mécanique de passation des différents niveaux associés à chacune des SE. Nous décrirons le processus de validation que nous avons mis en branle avant de procéder à la collecte des données. Nous enchainerons en décrivant notre échantillon expérimental. Enfin, nous préciserons les procédures que nous avons suivies pour traiter les données.

Les situations d'évaluation

Comme le soulignent à juste titre Laurier, Tousignant et Morissette (2005), la construction d'une situation d'évaluation est une étape cruciale pour s'assurer de recueillir une information utile concernant le niveau de développement de la compétence :

> Demander à des élèves, sans trop fixer de balises, de réaliser un projet sur des sujets comme la pollution atmosphérique, un programme communautaire ou une innovation technologique ne suffit pas pour obtenir des informations utiles pour l'évaluation. Il faut être précis dans la conception de la tâche et dans sa présentation aux élèves. (p. 109)

C'est pourquoi nous avons ciblé la première compétence du programme de science et technologie qui s'énonce comme suit : « chercher des réponses ou des solutions à des problèmes d'ordre scientifique ou technologique » (MELS, 2003c). En ce qui nous concerne, nous avons choisi – pour des raisons liées à l'accès aux sujets – de traiter uniquement la partie d'investigation scientifique en laboratoire dans le cadre de notre expérimentation.

L'architecture des situations d'évaluation s'inspire de la proposition de Rey, Carette, Defrance et Khan (2003) qui ont proposé initialement le modèle, mais dans d'autres disciplines (mathématiques et français) et dans une perspective d'évaluation diagnostique. En ce qui nous concerne, nous avons repris l'idée à la base de leurs réflexions et nous l'avons adaptée au contexte québécois de l'enseignement scientifique. Ainsi, chacune des situations d'évaluation se présente en trois « versions » qui correspondent à des niveaux. Au niveau 1 (N1), le plus avancé, il s'agit d'une situation d'évaluation qui se limite principalement à l'énoncé du problème à résoudre. Il n'y a aucun indice quant à la solution ou à la méthode à employer pour résoudre ce problème. Ce niveau

correspond au niveau de maîtrise « compétence assurée ». Au niveau 2 (N2), il s'agit du même problème, accompagné d'indices essentiellement conceptuels et méthodologiques qui ont pour but d'accompagner l'élève dans la résolution de la tâche qu'il doit accomplir. Ce niveau correspond, quant à lui, au niveau de maîtrise « compétence partielle ». Au niveau 3 (N3), on présente toujours la même situation, mais cette fois en guidant l'élève vers une solution unique et une méthode permettant d'y arriver. Il ne s'agit alors plus, à strictement parler, d'une situation visant à mesurer le niveau d'atteinte d'une compétence mais plutôt d'une situation qui permet de vérifier la maîtrise de certaines habiletés scientifiques (ex. construire un graphique, énoncer une hypothèse, etc.). L'administration des situations d'évaluation se réalise selon la séquence illustrée à la figure 1.

Afin de tenir compte du contexte interdisciplinaire, nous avons rédigé trois situations d'évaluation (SE) associées à trois contextes disciplinaires différents à savoir un contexte lié à la physique (la tasse de chocolat chaud), un contexte lié à la mise en œuvre d'une démarche dans laquelle il n'y avait pas de contexte disciplinaire prédominant (la balle de tennis) et un contexte associé à la biologie (les lombrics). En tout, il y avait donc neuf situations d'évaluation (3 SE X 3 niveaux/SE) qui sont résumées au tableau 1 qui suit.

Chacune des situations d'évaluation visait à rendre compte de la compétence d'investigation scientifique en laboratoire c'est-à-dire la compétence 1 du programme ministériel. En résumé, la SE1 consistait pour les élèves à élaborer une démarche d'investigation visant à choisir un contenant de remplacement qui soit tout aussi efficace que les tasses de polystyrène et plus écologique. En ce qui concerne la SE2, il s'agissait pour les élèves de choisir une variable qui influence le rebond d'une balle de tennis et d'en mesurer les effets. Enfin, la SE3 demandait aux élèves de choisir une variable qui influence l'abondance et la vigueur des lombrics et d'en démontrer l'influence.

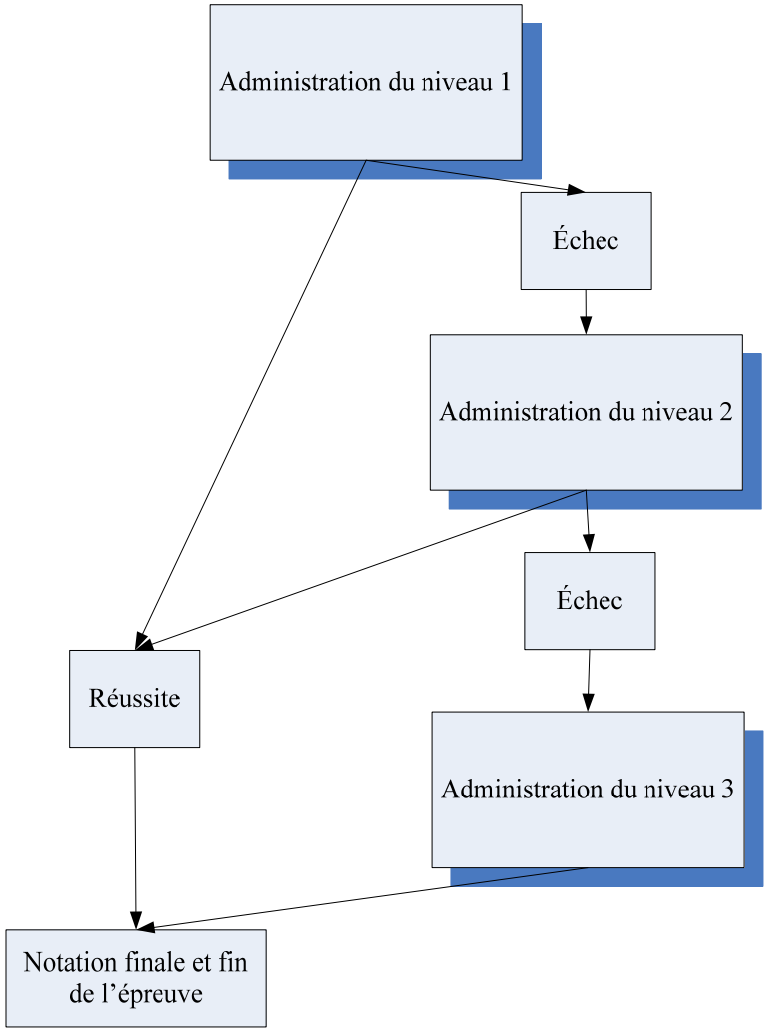

Figure 1
Séquence de passation des différents niveaux pour une situation d'évaluation.

Tableau 1
Description de chacune des situations d'évaluation par niveau.

	Niveau 1	Niveau 2	Niveau 3
SE1 (Physique) La tasse de chocolat chaud	SE1-1	SE1-2	SE1-3
SE2 (Démarche) La balle de tennis	SE2-1	SE2-2	SE2-3
SE3 (Biologie) Les lombrics	SE3-1	SE3-2	SE3-3

La validation des situations d'évaluation

Lors du processus de validation, nous avons soumis nos épreuves au jugement d'un comité de cinq experts composé de didacticiens des sciences, de spécialistes en mesure et évaluation et de praticiens chevronnés. Par la suite, nous nous sommes inspirés d'une proposition de Solano-Flores & Shavelson (1997) et nous avons réalisé une première passation avec un échantillon de 102 élèves issus de quatre groupes-classes. Cette passation a permis de réviser et de bonifier les instruments en vue de la collecte des données.

L'échantillon expérimental

Les situations d'évaluation validées ont ensuite été administrées à notre échantillon expérimental composé de 560 élèves du premier cycle du secondaire issus de 49 groupes-classes. Au total, 22 enseignants en provenance de cinq commissions scolaires et de 14 écoles ont accepté d'administrer les épreuves. Le tableau 2 illustre la répartition des élèves de notre échantillon selon le sexe et l'âge. La majorité des élèves se situaient dans un parcours standard et ne manifestaient pas de retard scolaire alors qu'une minorité (6,6%) avait au moins une année de retard.

Tableau 2
Distribution des effectifs de l'échantillon selon le sexe et l'âge.

Caractéristique	Effectif	Pourcentage
Genre		
Garçons	247	44,1
Filles	313	55,9
Âge (années)		
12	165	29,5
13	360	64,3
14	31	5,5
15	4	0,7

Afin de standardiser les conditions de passation des situations d'évaluation, des séances de formation à l'intention des enseignants et aussi des techniciens de laboratoire ont eu lieu. On y a alors présenté les différents documents expliquant la démarche à suivre avant, pendant et après la passation des SE. Un site web a été développé pour aiguiller les enseignants durant le parcours d'administration des épreuves et des films explicatifs sous format DVD ont aussi été mis à la disposition des enseignants participant à cette recherche.

La cueillette et le traitement des données

Chaque élève avait pour mandat de réaliser l'une des trois situations d'évaluation. Ces dernières ont été distribuées au hasard par l'enseignant à chacun des élèves. Le temps consacré à la réalisation des épreuves variait de 4 à 6 périodes de 75 minutes ce qui représente un temps de présence en classe entre 5 et 7 heures. Chacune des copies a ensuite été corrigée à l'aide d'une grille de notation comportant 12 critères qui représentaient des indicateurs pertinents de la compétence d'investigation scientifique. Le tableau 3 présente les critères de correction que nous avons retenus ainsi que la correspondance entre ces derniers et les composantes de la compétence d'investigation scientifique décrite dans le programme du MELS (Ministère de l'Éducation du Québec, 2003c, p. 277). Lors de la notation, nous avons contre-vérifié 10% des notes attribuées afin de nous assurer de la constance du jugement.

Tableau 3
Critères de notation retenus pour la correction des situations d'évaluation et correspondance avec les composantes de la compétence d'investigation scientifique du MELS.

Critères	Composantes
1. Reformulation du problème 2. Identification de la variable à investiguer 3. Élaboration d'une liste de matériel	1. Cerner un problème
4. Formulation d'une hypothèse vérifiable et plausible 5. Formulation et présentation des manipulations	2. Choisir un scénario d'investigation
6. Consignation des données expérimentales et des observations 7. Identification des causes d'erreurs 8. Traitement des données expérimentales	3. Concrétiser sa démarche
9. Maîtrise des concepts, des règles et des techniques 10. Application des résultats expérimentaux dans la solution propose	4. Analyser ses résultats ou sa solution
11. Présentation générale 12. Utilisation des symboles et des termes	5. Communiquer à l'aide des langages utilisés en science[5]

Les résultats ont ensuite été traités statistiquement avec la version 15 du logiciel SPSS (2006) en ayant recours essentiellement à des statistiques non paramétriques. Pour comparer les moyennes, nous avons utilisé le test U de Mann-Whitney pour deux échantillons indépendants et l'analyse de variance à un critère de classification de Kruskal-Wallis (Howell, 1998). Ce choix se justifie par le fait que nos données ne possédaient pas toujours des distributions qui satisfaisaient les condi-

[5] Cet énoncé n'est pas une composante de la compétence d'investigation scientifique mais plutôt l'énoncé de la compétence de communication scientifique.

tions d'utilisation des tests paramétriques. Compte tenu de ce contexte, nous avons fait le choix de demeurer prudent et d'utiliser des tests qui sont moins contraignants relativement à la distribution des données.

RÉSULTATS

Les résultats concernant la première question de recherche (Est-ce que le modèle à trois niveaux que nous proposons permet de juger du développement d'une compétence d'investigation scientifique ?) seront présentés en trois parties. Premièrement, nous dresserons un portrait des différents profils de passation que nous avons obtenus. Deuxièmement, nous montrerons les résultats obtenus par les élèves de chacun de ces profils pour chacune des situations d'évaluation. Troisièmement, nous jetterons un coup d'œil sur les taux de réussite pour les trois situations d'évaluation que nous avons expérimentées. Nous terminerons en présentant les résultats concernant la seconde question de recherche (Est-ce que le modèle se comporte de la même façon selon le contexte disciplinaire?) en abordant les spécificités disciplinaires.

Les profils de passation

Le tableau 4 présente chacun des profils avec les effectifs correspondants. D'abord, un nombre pratiquement égal d'élèves (189, 186 et 185) se sont vus offrir chacune des situations d'évaluation ce qui n'est pas surprenant puisque les SE étaient attribuées au hasard.

Tableau 4
Distribution des élèves pour chacun des profils de passation.

SE	Niveaux					
	1 N (%) (Profil 1)	1-2 N (%) (Profil 2)	1-2-3 N (%) (Profil 3)	1-3 N (%) (Profil 4)	3 N (%)	Total N (%)
1	92 (16,4%)	19 (3,4%)	24 (4,3%)	54 (9,6%)	0 (0,0%)	189 (33,8%)
2	77 (13,8%)	10 (1,8%)	59 (10,5%)	40 (7,1%)	0 (0,0%)	186 (33,2%)
3	83 (14,8%)	4 (0,7%)	31 (5,5%)	66 (11,8%)	1 (0,2%)	185 (33,0%)
Total N (%)	252 (45,0%)	33 (5,9%)	114 (20,4%)	160 (28,6%)	1 (0,2%)	560 (100,0%)

On constate également qu'il se dégage quatre profils d'élèves distincts. Le profil 1 regroupe les élèves qui ont eu besoin uniquement du niveau 1. En d'autres termes, ce sont des élèves qui manifestent une maîtrise jugée adéquate de la compétence d'investigation scientifique. Le profil 2 représente les élèves qui après avoir manifesté des difficultés à résoudre la situation du niveau 1 ont réussi après s'être vus administrer le niveau 2. Les élèves du profil 3 étaient en difficulté aux niveaux 1 et 2 et ils ont donc eu besoin des consignes intégrées au niveau 3. Un quatrième profil, inattendu, est apparu lors de l'analyse des données et représente les élèves qui ont réalisé le niveau 1 et le niveau 3 sans passer par le niveau 2. Après avoir vérifié auprès des enseignants concernés, ces derniers nous ont dit avoir fait un accroc au protocole pour deux raisons. Certains nous ont affirmé vouloir vérifier si les élèves qui réussissaient la situation de compétence assurée (niveau 1) réussissaient aussi la situation axée sur la simple mesure des habiletés (niveau 3). Cette dernière

correspond, selon les enseignants interrogés, aux situations habituelles auxquelles ils se réfèrent traditionnellement. D'autres nous ont mentionné que la situation de niveau 3 a été administrée à des élèves qui avaient réussi le niveau 1 afin de les occuper pendant la ou les période(s) restante(s).

Résultats pour la situation d'évaluation 1 « La tasse de chocolat chaud » par profil de passation

La figure 2 illustre la distribution des scores des élèves pour la situation d'évaluation 1 intitulée « la tasse de chocolat chaud » et dont le contexte disciplinaire principal est lié à la physique. Les résultats apparaissent pour chacun des quatre profils de passation. Sur l'échelle de notation, le point de césure a été fixé à l'échelon 2. Le trait horizontal illustre ce point de césure. Un coup d'œil général sur la figure 2 nous indique que le modèle semble cohérent lors du passage d'un niveau à l'autre. Ainsi, les élèves du profil 1 ont des résultats qui montrent une maîtrise suffisante de la compétence avec un score moyen de 2,24 qui se situe au-delà du point de césure. On peut facilement remarquer que les élèves du profil 2 sont clairement en difficulté face au niveau 1 de cette SE. Le score moyen qu'ils obtiennent se situe en effet à 0,36.

Cependant, le niveau 2 leur permet de faire passer ce score moyen à 1,72 ce qui dénote une intéressante amélioration. Malgré tout, ce score se situe en-deçà du score minimal attendu. L'analyse des scores du profil 3 montre une tendance similaire. Les élèves de ce profil sont en grande difficulté aux niveaux 1 et 2 avec des scores moyens respectifs de 0,49 et 0,72. Pour ces élèves, l'administration du niveau 3 est bénéfique car les résultats indiquent que malgré une maîtrise insuffisante de la compétence, ils semblent malgré tout maîtriser les habiletés inhérentes au succès de la situation d'évaluation. De fait, le score moyen de ces élèves passe à 2,08 lorsqu'on leur administre la situation dirigée du niveau 3. Les distributions associées aux élèves du profil 4 montrent peu de différences entre elles. La différence entre les scores moyens (2,46 et 2,78) est somme toute statistiquement significative.

Dans le cas de la première situation d'évaluation, on constate que les scores moyens (2,73) de la situation de niveau 3 sont légèrement plus élevés que pour la situation de niveau 1 (2,46). En jetant un coup d'œil

global sur la figure 2, on peut remarquer une tendance générale à la hausse dans la distribution des scores quand on passe d'un niveau à l'autre.

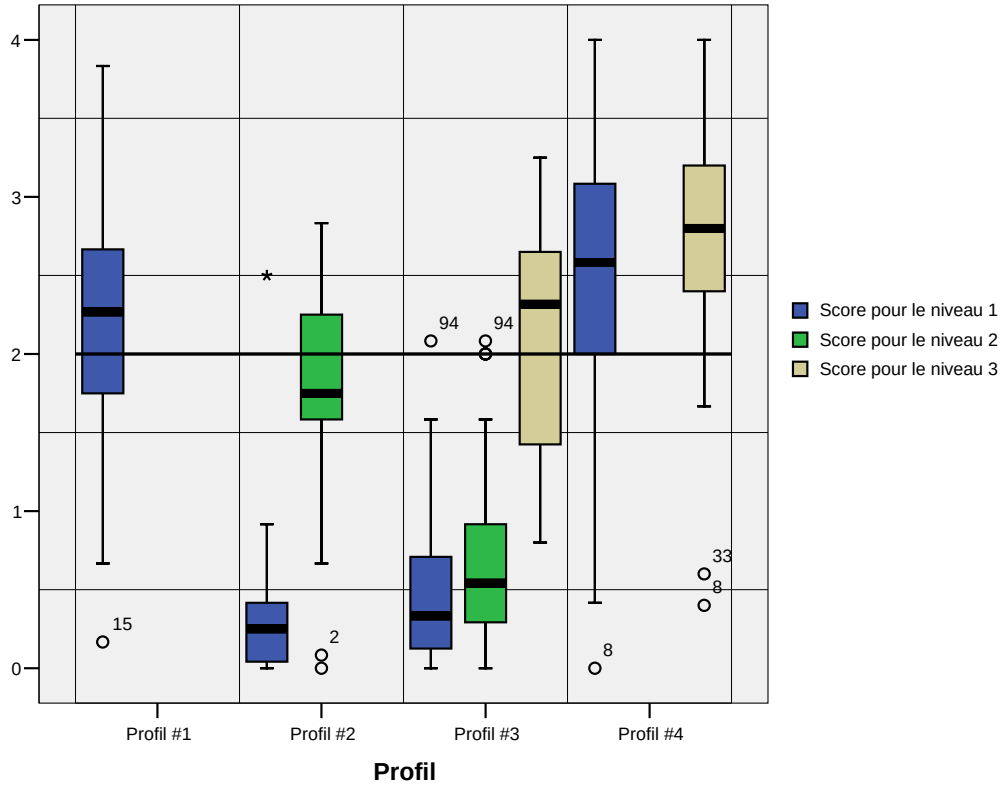

Figure 2
Distribution des scores pour la situation d'évaluation 1 concernant les niveaux 1, 2 et 3 pour chacun des quatre profils de passation.

Résultats pour la situation d'évaluation 2 « La balle de tennis » par profil de passation

Les distributions des scores pour la situation d'évaluation 2 sont présentées à la figure 3 pour chacun des quatre profils de passation. Les élèves appartenant au profil 1 ont obtenu un score moyen de 2,14 ce qui se situe juste au-dessus du score de césure.

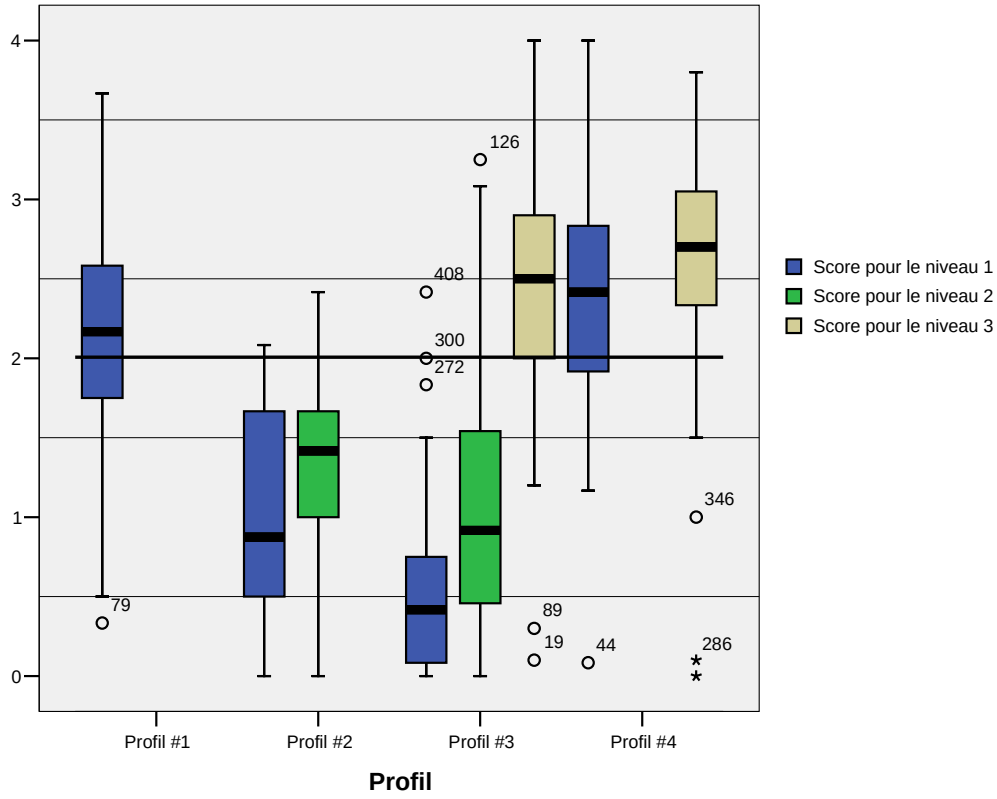

Figure 3
Distribution des scores pour la situation d'évaluation 2 concernant les niveaux 1, 2 et 3 par profil de passation.

Les distributions associées aux scores des élèves du profil 2 nous indiquent que la différence entre les scores moyens (N1 : 0,98, N2 : 1,36), bien que significative, demeure faible. Le fait d'administrer le niveau 2, autrement dit de donner des indices aux élèves, ne les aide pas vraiment à atteindre le seuil de passage. En ce qui concerne les scores des élèves du profil 3, on remarque à nouveau que les résultats pour les deux premiers niveaux sont très faibles. En effet, les scores moyens se situent respectivement à 0,51 et 1,10. Pour ces élèves, le niveau 3 semble salutaire, car le score moyen se situe bien au-dessus du point de césure à 2,44. Pour les élèves appartenant au profil 4, les scores moyens obtenus au niveau 1 et au niveau 3 sont relativement élevés à 2,35 et 2,57. Le test de la somme des rangs de Wilcoxon indique que cette différence est significative.

Résultats pour la situation d'évaluation 3 « Les lombrics » par profil de passation

La figure 4 présente les différentes distributions des scores pour la troisième situation d'évaluation que nous avons expérimentée. L'allure générale des distributions représentées à la figure 4 montre des ressemblances manifestes avec les distributions que nous avons jusqu'à maintenant décrites.

Cette fois, on remarque que les élèves du profil 1 ont obtenu un score moyen se situant tout juste au-dessus du point de césure à 2,04. Les élèves du profil 2 ont été constamment en difficulté avec des scores moyens s'élevant respectivement à 0,25 pour le niveau 1 et 1,27 pour le niveau 2. Les distributions des élèves du profil 3 montrent, quant à elles, que les élèves ont obtenu des résultats très faibles pour les deux premiers niveaux (N1 : 0,54 et N2 : 0,95). Le recours au niveau 3 semble avoir été pour eux bénéfique car le score moyen s'est alors élevé à 2,06 ce qui est similaire au score moyen obtenu par les élèves du profil 1. Enfin, les élèves du profil 4 ont obtenu des scores moyens satisfaisants c'est-à-dire se situant au-dessus du point de césure. Au niveau 1, le score moyen est de 2,07 et pour le niveau 3 il se situe à 2,38. Cette différence n'est pas significative.

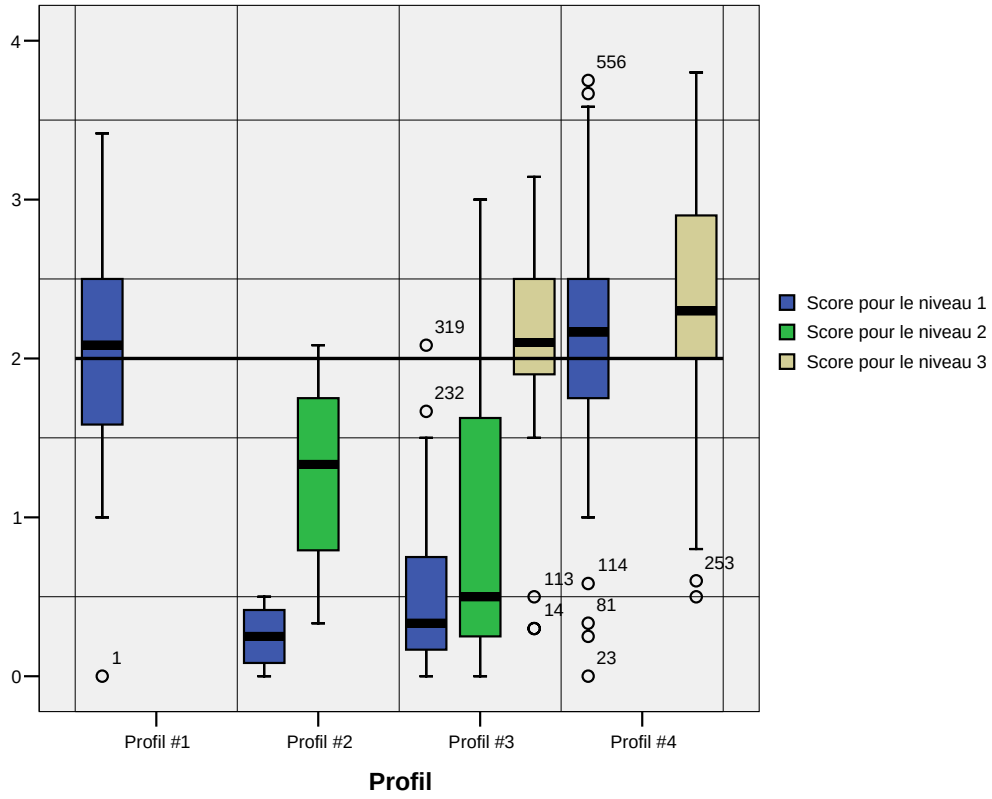

Figure 4
Distribution des scores pour la situation d'évaluation 3 concernant les niveaux 1, 2 et 3 par profil de passation.

L'effet des différents niveaux pour les élèves en grande difficulté

L'intérêt de notre modèle consiste à offrir deux solutions de rechange (niveau 2 et niveau 3) aux élèves qui manifestent des difficultés importantes lorsque confrontés à une situation où ils doivent manifester leur compétence ce qui représente une situation difficile pour bien des élèves. Les résultats consignés dans le tableau 5 sont, à ce sujet, éloquents pour les cas extrêmes.

Tableau 5
Pourcentage des élèves ayant obtenu un score nul en regard des 12 critères de notation.

Niveaux	Situations d'évaluation		
	1	2	3
1	5,29%	8,06%	4,89%
2	6,98%	2,94%	11,43%
3	0,00%	1,16%	0,00%

On remarque qu'au niveau 1, une proportion relativement importante (5,29%, 8,06%, 4,89%) d'élèves manifeste ce que nous qualifions de « syndrome de la page blanche » et que nous décrivions précédemment. Ces élèves, face à l'ampleur de la tâche qu'ils doivent accomplir, n'arrivent pas à manifester le moindre signe observable quant à la compétence qu'ils doivent démontrer. Pour un enseignant, cette situation est catastrophique, car il n'y a aucune trace sur laquelle s'appuyer pour étayer le jugement. Au niveau 2, on note même une augmentation des proportions pour les situations 1 et 3. Par contre, au niveau 3 il n'y a pratiquement plus d'élèves qui rencontrent suffisamment de difficultés pour remettre un cahier vide dans lequel aucune trace des apprentissages n'est visible.

Dans un modèle traditionnel à passation unique, on retrouverait effectivement uniquement le niveau 1 puisque le but de l'évaluation, si on veut respecter la cohérence curriculum-évaluation, doit permettre de circonscrire le niveau d'atteinte de la compétence sans fournir aucune aide et sans mesurer de simples habiletés.

L'effet du contexte disciplinaire

Le tableau 6 présente une synthèse des résultats à la suite de la comparaison des données des situations d'évaluation entre elles. On constate qu'il n'y a pas de différences significatives dans les moyennes obtenues pour les élèves du profil 1. En ce sens, le contexte disciplinaire ne semble pas jouer un rôle déterminant. Pour les résultats des élèves du profil 2, il n'a pas été possible de comparer statistiquement les données en raison de la taille des échantillons jugée trop faible. Les données laissent supposer que la situation 2 pourrait être un peu plus facile pour les élèves de

cette catégorie. En ce qui concerne les élèves du profil 3, les seules différences significatives concernent le niveau 3. La situation impliquant les balles de tennis semble alors plus facile pour les élèves. Enfin, les différences les plus importantes impliquent les élèves du profil 4. Pour ces élèves, la situation 1 est plus facile alors que la troisième semble être la plus difficile.

Tableau 6
Scores moyens par profil de passation pour chacun des niveaux de chacune des situations d'évaluation.

		SE1			SE2			SE3		
		N1	N2	N3	N1	N2	N3	N1	N2	N3
Profil 1	Moy.	2,24			2,14			2,04		
	é.t.[a.]	0,66			0,71			0,62		
Profil 2	Moy.	0,36	1,72		0,98	1,36		0,25	1,27	
	é.t.	0,57	0,80		0,74	0,68		0,22	0,72	
Profil 3	Moy.	0,49	0,72	2,08*	0,51	1,10	2,44*	0,54	0,95	2,06*
	é.t.	0,52	0,63	0,80	0,54	0,80	0,79	0,57	0,88	0,69
Profil 4	Moy.	2,46*		2,78*	2,35*		2,57*	2,07*		2,38*
	é.t.	0,91		0,77	0,78		0,84	0,74		0,71

a. é.t. : écart-type. * Résultats significatifs ($p \leq 0,05$).

DISCUSSION ET CONCLUSION

Les résultats que nous avons obtenus indiquent que le modèle d'évaluation que nous avons proposé permet de recueillir des informations auprès d'élèves ayant des niveaux fort différents de développement de la compétence. De fait, le modèle permet de catégoriser les élèves selon qu'ils : (1) maîtrisent la compétence de façon autonome, (2)

maîtrisent la compétence mais avec de l'aide ou (3) maîtrisent les habiletés sans être en mesure de les mobiliser en contexte. Puisque les élèves manifestent des degrés divers d'appropriation de la compétence, un modèle à plusieurs niveaux apparaît intéressant afin d'enrichir les données recueillies auprès des élèves. En effet, les résultats concernant uniquement le niveau 1 – la seule situation mesurant formellement la compétence – indiquent que le seul recours à cet outil ferait en sorte qu'il ne serait pas possible de dire quoi que ce soit sur le niveau de développement de la compétence avec aide ou sur le niveau de maîtrise des habiletés. Dans un contexte de certification où il faut amasser des éléments tangibles de preuves pour appuyer son jugement, cela représente un écueil de taille.

Notre modèle s'est inspiré, entre autres, des travaux de Rey *et al.* (2003) quant à la structuration des situations d'évaluation. Nos résultats concordent avec ceux qu'ils ont obtenus. On retrouve la même cohérence entre les niveaux. Ainsi, les élèves qui reçoivent une forme d'aide supplémentaire arrivent à mieux faire la démonstration de ce qu'ils sont en mesure de réaliser. Par contre, nous avons constaté que le niveau 2, et ce pour les trois situations d'évaluation, ne permettait d'améliorer que timidement la proportion d'élèves en mesure de réaliser la tâche qui leur était confiée. De toute évidence, il semblerait que l'aide fournie ne soit pas aussi bénéfique que prévu. Cette aide n'était peut-être pas suffisamment précise et explicite. Par exemple, afin d'aider les élèves à cerner le problème nous leur demandions : « Est-ce que tu as pensé à reformuler le mandat dans tes mots ? », « Est-ce que tu as pensé à identifier quelle variable tu mesurerais ? ». Pour des jeunes qui sont au premier cycle du secondaire, ces questions n'étaient peut-être pas aussi adaptées que nous le pensions. D'autres expérimentations seront cependant nécessaires afin de pouvoir mieux comprendre comment doser l'aide à fournir aux élèves afin qu'ils puissent faire la démonstration de leurs compétences.

De façon étonnante, nous n'avons pas remarqué de différences importantes liées au contexte disciplinaire. Les résultats sont comparables et le fait d'administrer une situation dans un contexte en particulier ne semble pas avoir d'effets sur la réussite des élèves. Il faut cependant mentionner que nos données sont limitées et qu'il est difficile de conclure sur l'effet du contexte disciplinaire. Cependant, nos résultats sont

particulièrement intéressants puisque, si d'éventuelles recherches confirmatoires venaient appuyer ce résultat, cela voudrait dire que le nombre de situations d'évaluation à administrer en fin de cycle pour sanctionner les élèves pourrait être grandement limité. Aussi, cela pourrait confirmer la pertinence d'une approche intégrée en science. Dans le cas inverse, il faudrait alors soit augmenter et diversifier les situations d'évaluation ou créer des règles de sanction qui viendraient identifier et limiter les contextes disciplinaires utilisés dans les épreuves certificatives.

RÉFÉRENCES

American Association for the Advancement of Science. (1993). *Benchmarks for science literacy, project 2061*. New-York: Oxford University Press.

Barma, S., & Guilbert, L. (2005). Le nouveau programme de formation en science et technologie: Fondements du programme et défis pour les enseignants. In *Sciences et technologies : de l'enseignement général aux contextes spécifiques d'adaptation,* ACFAS Annual Conference, Chicoutimi, Canada.

Beckers, J. (2002). *Développer et évaluer des compétences à l'école: Vers plus d'efficacité et d'équité*. Bruxelles: Éditions Labor.

Boutin, G., & Julien, L. (2000). *L'obsession des compétences*. Montréal: Éditions Nouvelles.

Brown, J. H., & Shavelson, R. J. (1996). *Assessing hands-on science: A teacher's guide to performance assessment*. Thousan Oaks, CA: Corwin Press.

Dionne, E. (2000). État de l'application du constructivisme en sciences physiques 416-436 dans une commission scolaire québécoise. Unpublished master's thesis, Université de Montréal, Montréal, Québec, Canada.

Dionne, E. (2005). Enjeux en évaluation des apprentissages dans le cadre de programmes d'études formulés par compétences: L'exemple de science et technologie. *Mesure et évaluation en éducation, 28*(2), 49-66.

Dionne, E. (2007). Un modèle d'évaluation certificative multiniveaux: Une voie prometteuse pour l'évaluation des compétences scientifiques en milieu scolaire ? *Dire, 16*(2), 26-29.

Dionne, E. (2008). *Expérimentation d'un modèle d'évaluation permettant de juger du développement d'une compétence d'investigation scientifique en laboratoire.*

Thèse de doctorat non publiée, Université de Montréal, Montréal, Québec, Canada.

Fourez, G. (1994). *Alphabétisation scientifique et technique: Essai sur la finalité de l'enseignement des sciences*. Bruxelles: De Boeck.

Francoeur, E. (2005). *Équiper les enseignants du secondaire pour l'enseignement des sciences et technologies: L'expérience de l'école de technologie supérieure.* In *Constats, défis et propositions en enseignement de la science et de la technologie: Quelles avenues à privilégier d'ici l'an 2015 ?*ACFAS Annual Conference, Chicoutimi, Canada.

Howell, D. C. (1998). *Méthodes statistiques en sciences humaines* (M. Rogier, V. Yzerbyt, & Y. Bestgen, Trans.). Bruxelles: DeBoeck. (Original work published 1997).

Jonnaert, P. (2002). *Compétences et socioconstructivisme: Un cadre théorique*. Bruxelles: DeBoeck.

Klein, S. P., Jovanovic, J., Stecher, B. M., McCaffrey, D., Shavelson, R. J., Haertel, E., et al. (1997). Gender and racial/ethnic differences on performance assessments in science. *Educational Evaluation and Policy Analysis, 19*(2), 83-97.

Laurier, M. D., Tousignant, R., & Morissette, D. (2005). *Les principes de la mesure et de l'évaluation des apprentissages* (3rd ed.). Montréal: Gaëtan Morin.

Legendre, R. (2005). *Dictionnaire actuel de l'éducation* (3rd ed.). Montréal: Guérin.

Leighton, J. P., Rogers, W. T., & Maguire, T. O. (1999). Assessment of student problem-solving on ill-defined tasks. *Alberta Journal of Educational Research, 45*(4), 409-427.

Mathy, P. (1997). *Donner du sens aux cours de sciences: Des outils pour la formation éthique et épistémologique des enseignants*. Bruxelles: DeBoeck.

Ministère de l'Éducation du Québec. (2001). *Programme de formation de l'école québécoise. Éducation préscolaire, enseignement primaire*. Québec: Gouvernement du Québec.

Ministère de l'Éducation du Québec. (2003a). *Plan de mise en oeuvre de la politique d'évaluation des apprentissages*. Québec: Gouvernement du Québec.

Ministère de l'Éducation du Québec. (2003b). *Politique d'évaluation des apprentissages: Formation générale des jeunes, formation générale des adultes, formation professionnelle*. Québec: Gouvernement du Québec.

Ministère de l'Éducation du Québec. (2003c). *Programme de formation de l'école québécoise. Enseignement secondaire, premier cycle*. Québec: Gouvernement du Québec.

Potvin, P., & Dionne, E. (2007). Realities and challenges of educational reform in the province of Québec: Exploratory research on teaching science and technology. *McGill Journal of Education, 42*(3), 393-410.

Rey, B., Carette, V., Defrance, A., & Kahn, S. (2003). *Les compétences à l'école: Apprentissage et évaluation*. Bruxelles: DeBoeck.

Scallon, G. (2004). *L'évaluation des apprentissages dans une approche par compétences*. Montréal: Éditions du renouveau pédagogique.

Solano-Flores, G., & Shavelson, R. J. (1997). Development of performance assessments in science: Conceptual, practical, and logistical issues. *Educational Measurement: Issues and Practice, 16*(3), 16-25.

SPSS inc. (2006). *SPSS 15.0 for Windows* [Computer software]. Chicago, IL: SPSS.

Tardif, J. (2006). *L'évaluation des compétences: Documenter le parcours de développement*. Montréal: Chenelière Éducation.

Eric Dionne est professeur à la Faculté d'éducation de l'Université d'Ottawa. Ses intérêts de recherche touchent à la modélisation des scores, à l'évaluation de programme ainsi qu'à l'évaluation scolaire et en particulier dans le domaine de l'éducation scientifique.

Michel D. Laurier est professeur titulaire à la Faculté des sciences de l'éducation de l'Université de Montréal. Il a été doyen de cette faculté de 2002 à 2010. Sa recherche porte sur l'évaluation des apprentissages et plus particulièrement sur l'évaluation des compétences langagières. Il est membre du GRIEMEtic, une équipe de recherche qui s'intéresse à l'utilisation des technologies en évaluation. Il est responsable de la dernière édition de l'ouvrage Principe de la mesure et de l'évaluation en éducation et a publié plusieurs articles dans le domaine de l'évaluation.