

Mesure et évaluation en éducation



Pratiques de pondération des résultats d'enquêtes à grande échelle aux notes scolaires en mathématiques Cas de l'Ontario français

Marielle Simon, Catherine Nadon, Christine van Barneveld, Sherry King et Gul Shahzad Sarwar

Volume 36, numéro 1, 2013

Date de réception : 27 février 2012

Date de réception de la version finale : 18 décembre 2012

Date d'acceptation : 20 décembre 2012

URI : <https://id.erudit.org/iderudit/1024463ar>

DOI : <https://doi.org/10.7202/1024463ar>

[Aller au sommaire du numéro](#)

Éditeur(s)

ADMEE-Canada - Université Laval

ISSN

0823-3993 (imprimé)

2368-2000 (numérique)

[Découvrir la revue](#)

Citer cet article

Simon, M., Nadon, C., van Barneveld, C., King, S. & Sarwar, G. S. (2013). Pratiques de pondération des résultats d'enquêtes à grande échelle aux notes scolaires en mathématiques : cas de l'Ontario français. *Mesure et évaluation en éducation*, 36(1), 1–25. <https://doi.org/10.7202/1024463ar>

Résumé de l'article

L'étude de nature interprétative rapportée dans cet article a pour but d'examiner les perceptions des membres du personnel enseignant à l'égard de la pratique de pondération d'une partie ou de l'ensemble des résultats des élèves à l'épreuve provinciale annuelle à leurs résultats de fin de cours en mathématiques 9^e année. En réponse à certains items du questionnaire contextuel à l'intention du personnel enseignant, administré en 2009-2012, 90% des enseignants des écoles de langue française de l'Ontario disent que cette pratique motive les élèves à prendre le test au sérieux. Leurs commentaires reflètent des préoccupations vis-à-vis des caractéristiques et comportements de l'élève et des conditions et caractéristiques du test, comme son enjeu et la taille de la pondération. Les constats rapportés rejoignent ceux proposés par Eklöf (2010) dans son modèle de motivation des élèves en situation d'évaluations à grande échelle.

Pratiques de pondération des résultats d'enquêtes à grande échelle aux notes scolaires en mathématiques : cas de l'Ontario français

Marielle Simon

Catherine Nadon
Université d'Ottawa

Christine van Barneveld
Université Lakehead

Sherry King
Université Lakehead

Gul Shahzad Sarwar
Université d'Ottawa

MOTS CLÉS: évaluation, évaluation à grande échelle, notation, motivation

L'étude de nature interprétative rapportée dans cet article a pour but d'examiner les perceptions des membres du personnel enseignant à l'égard de la pratique de pondération d'une partie ou de l'ensemble des résultats des élèves à l'épreuve provinciale annuelle à leurs résultats de fin de cours en mathématiques 9^e année. En réponse à certains items du questionnaire contextuel à l'intention du personnel enseignant, administré en 2009-2012, 90% des enseignants des écoles de langue française de l'Ontario disent que cette pratique motive les élèves à prendre le test au sérieux. Leurs commentaires reflètent des préoccupations vis-à-vis des caractéristiques et comportements de l'élève et des conditions et caractéristiques du test, comme son enjeu et la taille de la pondération. Les constats rapportés rejoignent ceux proposés par Eklöf (2010) dans son modèle de motivation des élèves en situation d'évaluations à grande échelle.

KEY WORDS: assessment, large-scale assessment, grading, motivation

In this qualitative study we examined teacher perceptions related to the optional practice of having part or all of a student's results to Ontario's annual Grade 9 province-wide assessment of Mathematics count for his or her final course grades. In response to items from the background questionnaire administered in 2009-

2012, 90% percent of teachers from Ontario's French-language schools agree that the practice motivates students to take the assessment more seriously. Related comments show concerns primarily regarding student characteristics and behavior as well as test conditions and characteristics such as the test stakes and the weight attributed to the results. The reported issues parallel those proposed by Eklöf (2010) in her model on the relationships between test stakes, test-taking motivation and test performance.

PALAVRAS-CHAVE: avaliação, avaliação em grande escala, notação, motivação

O estudo de natureza interpretativa apresentado neste artigo tem por finalidade examinar as percepções dos professores no que respeita à prática de ponderação de uma parte ou do conjunto dos resultados dos alunos no exame provincial anual com os seus resultados finais em matemática do 9.º ano. Em resposta a certos itens do questionário contextual, administrado em 2009-2012, 90% dos professores das escolas de língua francesa do Ontário dizem que esta prática motiva os alunos a levar o teste a sério. Os comentários refletem preocupações em relação às características e aos comportamentos dos alunos e às condições e às características do teste, como, por exemplo, o seu grau de dificuldade e o peso atribuído. As constatações apresentadas neste artigo vão ao encontro daquelas propostas por Eklöf (2010) no seu modelo de motivação dos alunos em situação de avaliação em grande escala.

Note des auteurs – À l'automne 2008, l'équipe de recherche composée des chercheurs et des membres de l'Office de la qualité et de la responsabilité en Ontario (OQRE) obtenait une subvention conjointe du Conseil de recherche en sciences humaines (CRSH). L'étude rapportée dans cet ouvrage porte sur une partie de cette étude. Toute correspondance peut être adressée comme suit : Marielle Simon, Professeure titulaire, Faculté d'éducation, Université d'Ottawa, téléphone : 613-562-5800 (2661), ou par courriel à l'adresse suivante : [msimon@uOttawa.ca].

Lors de la rédaction du manuscrit, Catherine Nadon et Gul Shazhad Sarwar terminaient leur programme de doctorat à l'Université d'Ottawa et Sherry King finissait son programme de maîtrise avec thèse à l'Université Lakehead.

Introduction

La présence d'enquêtes à grande ou à large échelle du rendement des élèves tend à s'accroître. Les résultats à ce type de tests sont souvent interprétés comme un miroir des compétences d'une population donnée à l'intérieur d'un secteur académique ou encore comme un indice phare de la qualité de l'éducation qui leur est dispensée. Les incidences de ces recherches à grand déploiement teignent inévitablement les débats en matière de politiques éducatives (Gustafsson, 2008). Les responsables en administration scolaire et des instances gouvernementales ainsi que les parents utilisent ces résultats dans le but de concevoir des plans d'aide à la réussite, d'allouer des ressources, d'orienter les prises de décisions et de mettre en place des politiques éducatives (Barker, 2004 ; Brown, 1999 ; Coburn & Talbert, 2006 ; Copland, 2003 ; Kerr, Marsh, Ikemoto, Darilek, & Barney, 2006 ; Kornhaber, 2004 ; Wayman & Stringfield, 2006a, 2006b ; Young, 2006). En dépit du sérieux accordé aux résultats de ces enquêtes à grande échelle, ces dernières ouvrent aujourd'hui la voie à des considérations plus larges concernant l'administration de ce type de test. La prise en compte du contexte d'administration de ce dernier vise à informer le public avec acuité sur les facteurs en mesure d'influencer le rendement des élèves interrogés afin de parfaire la validité et la représentativité des résultats de ces enquêtes.

Si le milieu de l'éducation accorde une grande importance à ces enquêtes à grande échelle, ce n'est pas autant le cas de la population étudiante qui ne voit pas toujours l'enjeu personnel lié à sa performance à celles-ci. Lorsque les membres de cette population croient que les divers éléments sur lesquels ils sont interrogés n'affecteront pas leurs résultats scolaires, la tendance populaire tient pour acquis que ces derniers, en général, ne mobilisent pas de manière optimale leurs compétences et leur effort pour répondre adéquatement aux questions et pour terminer le test selon leur plein potentiel.

Ce phénomène pointe vers un réel problème de validité des résultats puisque les enquêtes à grande échelle portent à croire qu'elles fournissent un portrait juste des capacités des élèves. Face à cette situation, un nombre croissant de systèmes scolaires et d'établissements d'enseignement adoptent des

politiques et des pratiques de pondération, totale ou en partie, des résultats de ces enquêtes aux résultats finaux des élèves à leurs cours de secondaire, afin de les encourager à remplir ces évaluations et à y fournir un véritable effort.

L'exemple du test de 9^e année en mathématiques dispensé par l'Office de la qualité et de la responsabilité en éducation (OQRE) en Ontario est un exemple éloquent de cette pratique. Inversement aux politiques de toutes les autres provinces et des territoires canadiens, l'Ontario n'oblige pas la pondération des résultats du test provincial dans le calcul de la note finale aux cours de mathématiques du secondaire. Selon les lignes directrices de l'OQRE, l'examen administré sur tout le territoire de la province ontarienne et à tous les élèves peut soit ne pas contribuer du tout à la note finale, soit ne représenter que jusqu'à 30 % des résultats du cours de mathématiques de 9^e année du secondaire.

Puisque ces conditions laissent toute une latitude dans la nature et l'étendue des décisions prises actuellement dans les divers milieux scolaires en Ontario en ce qui concerne les politiques de pondération, l'étude de cette pratique et de son rôle sur l'établissement de la validité des résultats devient particulièrement pertinente. La présente étude a donc pour but d'examiner les perceptions des membres du personnel enseignant vis-à-vis de cette politique et des pratiques courantes en province. Elle se base sur les données provenant des membres du personnel enseignant des écoles de langue française de l'Ontario qui ont participé à l'administration du test de mathématiques 9^e année lors de l'année scolaire 2009-2010. Étant donné le fait que les élèves de ces écoles obtiennent systématiquement des résultats inférieurs à ceux des élèves des écoles de langue anglaise aux enquêtes internationales, la prise de conscience des perceptions des membres du personnel enseignant de cette population à l'égard de la pratique de pondération à des fins incitatives constitue une étape essentielle dans l'étude de ce phénomène.

Contexte

Les enquêtes à grande échelle jouent actuellement un rôle de premier plan dans la recherche en éducation tout comme dans le milieu de l'éducation (voir Lafontaine & Simon, 2008 et Loye, 2011, pour un panorama des enquêtes internationales en éducation). Compte tenu de la nature de ces enquêtes et de l'importance accordée aux résultats, plusieurs mesures sont mises en place afin d'assurer la validité des données, notamment la standardisation de leurs

contenus, de leur format et de leur administration. Malgré les protocoles et l'appareillage déployés, certains biais interviennent et viennent brouiller les données relatives aux compétences et aux connaissances des élèves interrogés.

Dans plusieurs pays, États ou provinces, plusieurs politiques ont été mises sur pied récemment dans le but d'incorporer à leurs notes de fin de cours les résultats des élèves à certaines enquêtes à grande échelle, comme incitatif de motivation à passer ces tests. En Ontario, l'OQRE est responsable d'évaluer les apprentissages de tous les élèves de 3^e et de 6^e années de la province en mathématiques et en langue, les apprentissages en mathématiques de tous les élèves de 9^e année et les apprentissages en compétences linguistiques de tous les élèves de 10^e année. Le test provincial de mathématiques de 9^e année se distingue notamment de plusieurs enquêtes britanniques, américaines ou canadiennes puisque ses résultats n'ont pas d'incidence directe sur la diplomation de l'élève et puisque sa performance à cette évaluation n'est pas obligatoirement prise en compte dans son bulletin scolaire. Toutefois, les membres du personnel enseignant sont libres de le faire et tendent à en tenir compte en pratique (Koch, 2011).

Dans le contexte des enquêtes à grande échelle, la motivation de l'élève à s'impliquer pleinement dépend de ses attentes par rapport à son succès à l'épreuve et de la valeur qu'il accorde à cette épreuve (Eccles & Wigfield, 2002; Harlen & Crick, 2003; Van Barneveld, Pharand, Ruberto, & Haggarty, 2012; Wigfield & Eccles, 2000). Baumert et Demmrich (2001) définissent la motivation à passer une épreuve comme étant «la volonté de s'impliquer à répondre aux items de l'épreuve et d'investir effort et persistance à l'activité entreprise» (traduction libre, p. 441). La revue de la littérature révèle que l'approche de l'élève vis-à-vis des enquêtes à grande échelle influence significativement ses résultats (Drasgow, Levine, & Williams, 1985; Glas & Falcón, 2003; Glas & Meijer, 2003; Levine & Rubin, 1979; Nering, 1997; Van Krimpen-Stoop & Meijer, 1999). Un corpus grandissant de recherches décrit le phénomène à travers lequel l'élève ne mobilise pas toutes ses ressources et ne fournit pas nécessairement l'effort attendu pour remplir le test lorsque ce dernier n'a pas ou peu de conséquence personnelle (Baumer & Demmrich, 2001; Blau, Moller, & Jones, 2004; DeMars, 2000; Eklöf, 2010; Ferrer-Caja & Weiss, 2002; Hoyt, 2001; Olaru, Karageorghis, & Chatzisarantis, 2003; Sloane & Kelly, 2003; Van Barneveld et al., 2012; Wise, 1996a, 1996b; Wise & Demars, 2005; Wolf & Smith, 1995; Wolf, Smith, & Birnbaum, 1995). Autrement dit, il existe une relation entre les enjeux associés aux épreuves

tels que perçus par les élèves et leur motivation à compléter celles-ci (Eklöf, 2010) ; plus l'élève accorde de l'importance aux conséquences des épreuves, plus il est motivé à réussir (Ndinga & Frenette, 2010).

Plusieurs types d'incitatifs financiers et non financiers ont fait l'objet d'études par rapport à l'importance accordée aux conséquences des enquêtes à grande échelle telles que perçues par les élèves (Van Barneveld et al., 2012). La valeur de l'incitatif financier, le délai dans son attribution et le fait qu'il soit gagné ou perdu auraient une certaine incidence sur la motivation à terminer l'épreuve (Baumert & Demmrich, 2001 ; Levitt, List, Neckermann, & Sadoff, 2011 ; O'Neil, Abedi, Miyoshi, & Mastergeorge, 2005) quoique les conclusions demeurent partagées, surtout lorsque les études s'effectuent au palier secondaire.

Par ailleurs, certains incitatifs non financiers de nature politique (par ex., la mise sur pied de programmes d'études qui ont pour but d'aider les élèves à risque d'éviter de doubler leur année), académique (par ex., obtention d'un certificat, condition pour obtenir le diplôme d'études, comptabilisation des résultats dans la note finale) et pédagogique (par ex., rétroaction à l'appui de l'apprentissage) ont également fait l'objet d'études. Encore, les résultats sont mitigés et varient grandement selon les conditions méthodologiques (par ex., âge, niveau d'habileté, palier scolaire) (Levitt et al., 2011). Les constats brossés par ces études suggèrent que les résultats des enquêtes à grande échelle mènent à une sous-estimation des capacités réelles du rendement des élèves, surtout dans le cas des élèves à risque d'échec ou en difficulté d'apprentissage (Luce-Kapler & Klinger, 2005 ; Roderick & Engel, 2001 ; Van Barneveld et al., 2012).

Le peu de valeur accordé par les élèves aux enquêtes à grande échelle à faible enjeu peut se traduire par une propension de ces derniers à deviner les éléments de réponses, à omettre de répondre à certaines questions ou à refuser de remplir le test. Ces comportements s'observent davantage lorsque les résultats ne sont pas comptabilisés dans leurs résultats finaux (DeMars, 2000 ; Wolf & Smith, 1995), mais ils sont particulièrement prégnants quand les questions nécessitent une réponse à développement (DeMars, 2000 ; Gustafsson & Rosén, 2005), lorsqu'elles sont jugées comme étant difficiles (Wolf et al., 1995), quand elles exigent une lecture significative (Wise, 2006), quand elles apparaissent à la fin du test (Stocking, Steffen, & Eignor, 2001 ; Wise, 2006) ou encore lorsque des questions sont liées entre elles (Stocking et al., 2001).

Les écrits susmentionnés mettent en lumière plusieurs des facettes reliées à la motivation des élèves à remplir les épreuves à grande échelle et le besoin de questionner les personnes concernées, y compris les élèves et les membres du personnel enseignant. En effet, le personnel enseignant se préoccupe probablement davantage que les élèves de l'effort déployé par ces derniers lors des enquêtes, de leur préparation au test et de la possibilité d'abandonner, de deviner ou d'omettre les questions qui ne « comptent » pas aux fins du bulletin (Radford & Demers, 2010). Le personnel enseignant a également tendance à commenter les épreuves mêmes, notamment leur longueur, leurs fonctions, le moment de leur administration et plus directement la pratique de pondération des résultats de ces enquêtes aux fins des notes finales au cours.

L'intérêt à focaliser sur le personnel enseignant des écoles de langue française de l'Ontario découle des conditions démographiques qui les caractérisent et des résultats inférieurs qu'obtiennent les élèves qui les fréquentent lorsque comparés à ceux des élèves des écoles de langue anglaise. D'abord, les écoles de langue française de l'Ontario sont régies par un consortium des douze conseils scolaires de langue française qui en font partie. En plus de comprendre une grande diversité d'élèves sur le plan des compétences linguistiques, la situation minoritaire des écoles de langue française en Ontario a pour effet d'amener les élèves à s'identifier davantage à la culture anglo- que franco-canadienne, ce qui complique, dans une certaine mesure, la sélection de passages pertinents et captivants dans les enquêtes à grande échelle. Enfin, étant donné le faible bassin d'élèves dans ce milieu, ces derniers sont moins exposés que ceux de la cohorte anglophone aux tests standardisés en général et leurs écoles participent à un suréchantillonnage, ce qui risque de donner lieu à un désengagement des écoles, de leur personnel et du corps étudiant à l'ensemble des enquêtes à grande échelle.

Méthode

Au moment où l'OQRE a accepté de se joindre à l'étude actuelle, les responsables et les chercheurs ne connaissaient pas l'ampleur de la pratique en Ontario en raison de son statut optionnel. Quelques études anecdotiques (par ex., Koch, 2011) laissaient entendre qu'elle était répandue dans certains systèmes scolaires de langue anglaise. En partie aux fins de l'étude, les responsables de l'OQRE ont décidé d'ajouter sept nouvelles questions à la version 2010 du questionnaire à l'intention des membres du personnel enseignant en mathématiques, 9^e année (celles-ci sont fournies dans la section des résultats),

dont une les invitant à offrir un commentaire. Dans le présent ouvrage, un aperçu des résultats quantitatifs précède l'étude approfondie des commentaires fournis à la question # 22, point d'intérêt des chercheurs puisque les membres du personnel enseignant étaient complètement libres dans l'expression de leurs réponses. En effet, l'accent est mis sur l'étude de toutes leurs préoccupations émises, peu importe leur fréquence, le but étant d'en dégager une liste préliminaire tout en tenant compte du contexte particulier et des limites de l'étude.

Membres du personnel enseignant

Le test fut administré en deux sessions, c'est-à-dire à l'hiver 2010 à tous les élèves de 9^e année des cours appliqués ou théoriques de mathématiques d'automne 2009 et au printemps 2010 à tous les élèves des cours appliqués ou théoriques offerts à l'hiver 2010. Parallèlement, tous les membres du personnel enseignant responsables de ces cours devaient répondre à la version remaniée du questionnaire contextuel à leur intention. En Ontario français, le personnel enseignant de mathématiques, 9^e année, comptait 314 membres. Parmi ces derniers, 220 (70 %) ont répondu aux sept questions portant sur la motivation des élèves à remplir le test de l'OQRE. De plus, 33 % des répondants enseignaient principalement au programme théorique et 33 % possédaient entre zéro et cinq années d'expérience en enseignement des mathématiques, 9^e année. En ce qui concerne le nombre d'années d'expérience en enseignement des mathématiques, les quatre catégories offertes ont chacune été choisies par 25 % des répondants, notamment moins de deux ans d'expérience, entre trois et cinq ans, entre six et 10 ans et plus de 11 ans.

Analyse

À la suite de la compilation des réponses quantitatives aux sept questions portant sur la pratique de pondération, les commentaires obtenus à la question # 22 ont chacun été divisés en diverses unités de sens. L'utilisation d'une grille préliminaire formulée à l'aide des approches déductives (catégories issues de la recension) et inductives (catégories émergentes des données) a permis de faire ressortir les idées principales. Cette première version a servi à examiner un premier échantillon d'unités de sens provenant de l'ensemble des réponses des membres du personnel enseignant de toutes les écoles, tant de langue anglaise que de langue française, afin de remplir la grille, la taille de l'échantillon francophone étant trop faible. Un exercice de cotation entre juges a fourni un indice de concordance de 74 % (pour plus de détails, voir le

rapport rédigé par Van Barneveld, King, Simon, & Nadon, 2010). Cette analyse préliminaire a donné lieu à trois catégories : composantes reliées à l'élève, au test et aux valeurs.

En ce qui concerne les composantes reliées à l'élève, les commentaires peuvent se référer aux caractéristiques des élèves, comme sa disposition et sa motivation au test, son effort déployé, son niveau d'engagement, son niveau de stress ou de découragement ressentis lors du test et le programme d'études (ou type de cours) dans lequel les élèves sont inscrits. Les commentaires peuvent également porter sur le comportement de l'élève à l'égard du test. Ceux-ci comprennent les réactions observables de l'élève lors de l'administration du test comme l'omission volontaire de répondre à quelques questions ou à deviner certains éléments de réponses, le type de préparation au test, l'abandon et l'absence complète au test.

Par ailleurs, les composantes du test comprennent ses caractéristiques et les conditions d'administration. Les caractéristiques propres au test sont reliées à sa nature, à son format (par ex., le type de questions), à sa longueur (par ex., le nombre de questions) et aux composantes pondérées. La durée du test, la transmission des résultats, l'accès au test et les procédures de notation fournissent des exemples de conditions d'administration. Les commentaires peuvent également refléter des préoccupations à l'égard du moment où l'on informe la population enseignante de la date d'administration du test et du moment de l'année où le test est administré.

Enfin, la composante valeur est associée à la taille de la pondération attribuée aux résultats et à l'importance accordée par l'élève, par le personnel enseignant ou par d'autres membres de la communauté vis-à-vis du test. Tout autre commentaire qui ne propose qu'une simple reformulation de la question sans référence à d'autres éléments pertinents ou des thèmes non pertinents au cadre d'analyse est classé « autre ». Il va sans dire que la grille d'analyse a évolué et a été assujettie à des modifications à la suite des analyses des réponses de l'ensemble des membres du personnel enseignant aux deux sessions d'administration.

Résultats

Cette section présente les résultats aux sept questions à l'intention des membres du personnel enseignant mais cible davantage les commentaires fournis à la partie ouverte de la question # 22. Tous les pourcentages obtenus

aux questions de types fermés ont été calculés en fonction des 220 répondants. D'abord, à la question # 19 *Est-ce que vous attribuez des notes à certaines ou à toutes les composantes du Test de mathématiques, 9^e année, en vue de les inclure dans la note de cours de vos élèves?*, près de 90 % des répondants disent tenir compte de certaines ou de toutes les composantes du test dans leur calcul de la note de cours de leurs élèves. La taille de la pondération attribuée à ces composantes varie entre 6 à 15 % selon 50 % des répondants, et entre 26 à 30 % selon 26 % des participants. À la question # 21 *Avant de faire le Test de mathématiques, 9^e année, vos élèves ont-ils été informés du pourcentage que représenteraient ces notes dans leur note de cours (p.ex., 5%, 10% de la note obtenue en classe)?*, 90 % des répondants rapportent informer leurs élèves de la pratique. En réponse à la question # 23 *Quels types de questions notez-vous en vue de les inclure dans la note de cours de vos élèves?*, 38 % des répondants disent compter toutes les questions à réponses construites tandis que 48 % ne considèrent que quelques-unes de ces questions. Parallèlement, 54 % des répondants avouent noter toutes les questions à choix multiples tandis que 44 % ne tiennent compte que de quelques-unes d'entre elles. Les décisions relatives à cette pratique (question # 20 – d'appliquer ou non la pratique et question # 25 – quels items ou sections à pondérer) se prennent principalement sur le plan de l'école et en grande partie par les membres du personnel enseignant, de manière individuelle ou collective. Les résultats de la question # 24, qui demande à quel domaine (par ex., numération et algèbre) se rapportent les questions notées, se révèlent peu pertinents à l'étude.

En ce qui concerne les résultats de l'analyse des données de la question # 22 à composantes fermée et ouverte, 90 % des répondants sont d'accord avec l'énoncé qui dit que le fait d'inclure les résultats du test de l'OQRE dans la note de cours motive les élèves à prendre le test au sérieux. Les commentaires fournis par 49 (21 %) d'entre eux sont donc largement en faveur de la pratique. La déconstruction de leurs commentaires a donné 100 unités de sens. En fait, la plupart des commentaires ont fourni au moins deux unités de sens, jusqu'à un maximum de huit dans un cas particulier. Le tableau 1 présente la liste des unités de sens par thèmes. Les commentaires classés « autres » ont mené à la formulation de nouveaux sous-thèmes et le thème initial « valeur » a été fusionné aux thèmes « élève » et « test ». Organisés ainsi, les résultats reflètent une prédominance des considérations liées aux composantes de l'élève (50 %), suivi des composantes du test (40 %).

Tableau 1
Fréquence des unités de sens selon les thèmes et les codes

<i>Thèmes et codes</i>	<i>Nombre d'unités de sens par thèmes</i>
Considération liée aux composantes de l'élève	51
Caractéristiques	39
Indifférence (ou inverse - valorisation)	18
Effort	6
Motivation	6
Programme de l'étudiant	5
Habilités de l'élève	2
Stress	2
Comportements	12
Préparation	8
Réponse devinée	1
Omission à certains éléments du test	1
Copiage	1
Impossibilité de terminer l'ensemble du test	1
<i>Considération liée aux composantes du test</i>	40
Conditions entourant le test	24
Substitut pour l'évaluation sommative du cours	6
Renseignements fournis aux élèves	4
Moment de l'administration	4
Accessibilité au test	4
Exigences (p. ex., diplomation)	3
Moment de l'annonce de la date du test	2
Responsable des décisions relatives à la pratique	1
Caractéristiques du test	16
Durée	6
Format (questions à réponses choisies ou fermées)	5
Nature du test	2
Poids accordé au test	2
Longueur	1
<i>Autres</i>	9
Reformulation simple	5
Réponse floue	2
Texte illisible en raison de la mauvaise qualité de la numérisation	1
Réponse non pertinente	1
Total	100

Les commentaires des membres du personnel enseignant se référant à l'élève se regroupent en deux sous-catégories : les caractéristiques et les comportements. Les tableaux 2 et 3 présentent des commentaires types des caractéristiques et des comportements des élèves rapportés par le personnel enseignant. Les unités reliées aux caractéristiques des élèves portent plus spécifiquement sur la valeur accordée par les élèves au test, l'effort déployé, la motivation et les réalités propres aux programmes d'études dans lesquels sont inscrits les élèves. Les unités de sens se rapportant aux comportements des élèves portent sur leur préparation au test et sur les tendances à deviner, tricher, omettre des questions, abandonner le test ou de ne pas se présenter à l'épreuve.

Tableau 2
*Commentaires types du personnel enseignant
à l'égard des caractéristiques des élèves*

Caractéristiques de l'élève	Commentaires types du personnel enseignant
Valorisation du test	Je compte les notes du test comme l'examen final depuis plus de 7 ans et d'après les résultats des années précédentes, je conclus que les élèves prennent le testing très au sérieux. [8504039]
Effort	De façon générale, les élèves ont tendance à prendre au sérieux tout ce qui compte comme évaluation [8517036] Autrement il aurait moins d'effort [8516009] D'après mon expérience, les[sic] élèves ne mettent pas d'efforts au test s'il n'est pas évalué comme note au cours. [8517037]
Motivation	Je crois que ça n'affecte pas la motivation des élèves déjà motivés, mais aide aux élèves moins motivés à faire davantage d'effort. [8517022]
Programme d'études	Sans ces notes, l'élève au niveau appliqué s'en «ficherait[sic]»* du TPM [8502031]

(*opinion obtenue des élèves)

Tableau 3
*Commentaires types du personnel enseignant à l'égard
 des comportements de l'élève*

Comportements de l'élève	Commentaires types du personnel enseignant
Préparation	De cette façon ils étudient et se préparent[sic] comme tout[sic] autre évaluation, afin d'avoir la meilleure réussite possible. [8517031] [...] si ce n'est pas noté le TPM [test provincial de mathématiques] n'est pas pris au sérieux et ils ne se préparent pas en conséquence. [8502023]
Omission	[...] la moitié de la classe aurait [...] sauté les questions à développement. [8516051]
Réponse devinée	[...] la moitié de la classe aurait deviné les choix multiples. [8516051]
Copiage (tricherie)	[...] c'est une bonne stratégie pour éviter le copiage [...]. [8501033]
Impossibilité de compléter le test	Sauf quand les élèves ne peuvent pas terminer tout le cahier [...] [8516045] [...] et ne veulent pas que leurs résultats comptent car ils n'ont pas eu assez le temps pour réussir au niveau désiré. [8516057]

Le tableau 1 rapporte également des considérations liées aux composantes du test. Les unités de sens se regroupent en deux sous-catégories:

- a) conditions entourant le test et
- b) caractéristiques du test.

La première sous-catégorie regroupe des préoccupations par rapport au moment de l'annonce de l'administration du test, au moment de l'administration même du test et à l'accessibilité au test. D'autres unités de sens de cette sous-catégorie font ressortir des préoccupations reliées à l'idée de :

- a) renseigner l'élève par rapport à la pratique de pondération,
- b) substituer l'examen sommatif final du cours par le test de l'OQRE,
- c) rendre le test obligatoire aux fins de diplomation, comme dans le cas du Test provincial des compétences linguistiques (TPCL) de l'OQRE administré en 10^e et

- d) prendre des décisions liées à la pratique de pondération aux divers échelons administratifs.

Par ailleurs, des références à la durée, à la longueur, au format, à la nature et au poids accordé se rapportent au test lui-même. Les tableaux 4 et 5 présentent des commentaires types du personnel enseignant pour chacune de ces deux sous-catégories.

Tableau 4
Commentaires du personnel enseignant à l'égard des conditions entourant le test

Conditions entourant le test	Commentaires types du personnel enseignant
Substitut à l'examen du cours	[...] ils n'ont pas un deuxième examen à écrire après le PM (dans ce cours). [8517026] Pourquoi donner et préparer les élèves pour deux examens. [8517047]
Renseignements fournis aux élèves	Je dis que je les inclus mais je ne le fais pas. [8516051] Je dis aux élèves qu'une partie du test de l'OQRE sera complétée comme une partie de leur examen final [...] [8501033]
Moment de l'administration du test	Il est trop compliqué de le faire durant ce temps de l'année [...] [8516051]
Accès au test	[...] car on ne peut pas ouvrir les tests [...]. [8516051] [...] il faudrait qu'on puisse faire des copies [...] [8516030]
Exigences	Il faut que le TPM soit une exigence pour le DESO (à l'image du TPCL) [...]. [8502038]
Moment de l'annonce du test	[...] mes élèves et moi avons été informés du testing la journée avant le test. [8516001]
Personne responsable des décisions relatives à la pratique	[...] l'enseignant peut décider de la ou des questions à corriger (uniforme dans[sic] la province). [8504030]

Tableau 5
Commentaires du personnel enseignant à l'égard du test

Test	Commentaires du personnel enseignant
Durée	En 9 ^e , deux évaluations de 50 minutes, c'est suffisant. [8517047]
Format	Pour être capable de noter les questions construites[sic] [...]. [8516030]
Nature	[...] Que mesure ce TPM au juste? [...] [8504025]
Poids	Par contre, le test de l'OQRE devrait [...] représenter 30% de la note. [...] [8504030]
Longueur	[...] Il [le test] est long [...] [8504002]

Discussion

Dans son plus récent ouvrage, Eklöf (2010) propose une théorie de motivation à remplir un test. Le modèle relie les caractéristiques des élèves et du test qui jouent sur cette motivation et présente le lien entre cette dernière et la performance au test. Les facteurs associés à la motivation à remplir un test, retenus dans son modèle, proviennent d'une recension exhaustive des écrits et de ses études sur ce que rapportent les élèves à ce sujet. Dans cette section, les auteurs tentent de mettre en parallèle les thèmes issus des commentaires fournis par des membres du personnel enseignant de 9^e année en Ontario français et ceux du modèle d'Eklöf en ce qui concerne la pratique de la pondération de résultats de tests à grande échelle à la note finale au cours à des fins d'incitatif de motivation à les remplir et à les réussir.

Caractéristiques reliées aux élèves

Dans son modèle de motivation au test, Eklöf (2010) présente quatre caractéristiques associées aux élèves : leurs buts et orientations personnels, leurs attitudes générales à l'égard des tests, leur perception de soi et leur personnalité. Dans notre étude, l'examen des unités de sens issus des commentaires en fait ressortir trois : l'indifférence/valorisation qu'accorde l'élève au test en raison des conséquences qui y sont associées (ce qui se traduit en effort et motivation), son niveau d'habileté, défini principalement en termes du programme d'études de l'élève et le stress associé au test. Les commentaires mettent en évidence la croyance du principe du moindre effort tout en faisant ressortir la différence motivationnelle entre les élèves issus du programme théorique et ceux du programme appliqué, et plus spécifiquement les élèves

à risque. Selon les répondants, sans une mesure de pondération, les élèves à risque et ceux du programme appliqué rempliraient le test pour «s'en débarrasser», comme l'écrit nettement un membre. La problématique des élèves à risque rejoint ce qui se dit sur le plan international. Les politiques de sept États américains, par exemple, qui obligent la pratique de pondération des résultats des tests dans la note finale révèlent une très grande sensibilité envers les élèves à risque et les besoins d'adaptation et de remédiation en cas d'échec (*Center for Educational Policy*, 2010). Plus les enjeux des tests sont élevés plus les responsables tentent d'anticiper les possibilités de litiges en offrant des cours d'appoint, des séances de préparation et d'information et des consignes propres à toute une gamme d'adaptations potentielles. Enfin, les commentaires font référence au taux de stress que vivent les élèves peu importe leur programme ou leur niveau d'habileté général. Ils reflètent l'idée de stress en contexte général d'examen.

Caractéristiques reliées au test

Dans son modèle théorique, Eklöf (2010) propose également quatre caractéristiques reliées au test : son enjeu, le format des items (par ex., choix multiples), le contexte d'administration et la charge cognitive de ses items. Les quatre caractéristiques du test qui ressortent des unités de sens analysées dans notre étude, soit l'enjeu, le format des items, la nature du test et les conditions d'administration, se comparent à ceux du modèle. L'enjeu du test dépend de la taille de la pondération, de la fonction du test (par ex., exigences du programme en termes de sanction) et de la validité apparente du test aux yeux des élèves. Selon les propos des répondants, une pondération de 5 % à la note finale suffirait à motiver les élèves performants mais il faudrait un poids minimal de 30 % pour encourager les élèves des cours appliqués à remplir le test. Ceci explique la polarisation des résultats quantitatifs obtenus à la question # 19, au sujet de la taille accordée à la pondération. La variabilité dans les pondérations accordées aux diverses épreuves provinciales et territoriales au Canada et ailleurs témoigne d'un manque de consensus par rapport à la taille idéale de la pondération à considérer. Par exemple, les pondérations prescrites à travers le Canada varient de 10 %, à Île-du-Prince-Édouard, à 50 %, à Terre-Neuve et Labrador, Alberta, Nunavut, Territoires-du-Nord-Ouest et Québec (Volante, 2008). À ce sujet, les écrits n'ont pas encore confirmé une relation directe entre la taille de la pondération et la motivation à remplir le test. Par ailleurs, selon des répondants, l'absence d'une pondération peut mener les élèves à deviner ou à omettre des questions en raison d'un manque d'intérêt

et de valorisation envers le test. Inversement, une pondération élevée pourrait pousser certains élèves stressés à tricher au test afin d'assurer la réussite. Dans sa thèse de doctorat, Stone (2008) présente des commentaires d'élèves du palier secondaire qui décrivent explicitement leur tendance à copier et à tricher lors des épreuves nationales américaines, étant donné l'enjeu perçu. Ces constatations laissent supposer un besoin continu d'explorer davantage le lien entre la taille de la pondération et son impact sur la motivation, le comportement et la performance de l'élève aux épreuves dans le but de déterminer une pondération « idéale ».

En ce qui concerne la fonction du test comme deuxième dimension reliée à son enjeu, les unités de sens invitent à considérer la possibilité de remplacer complètement ou en partie l'examen du cours de mathématiques par le test provincial plutôt que d'exiger des élèves qu'ils soient soumis à deux ou plusieurs épreuves à la même période de l'année. Habituellement, les enquêtes à grande échelle sont administrées à la période des examens de fin de semestre ou d'année. La pratique de substitution des examens de fin d'années par une épreuve à grande échelle s'applique déjà dans toutes les autres provinces et les territoires du Canada (Klinger, DeLuca, & Miller, 2008 ; Volante, 2008) ainsi que dans au moins sept États américains et l'on anticipe que la pratique sera rendue obligatoire dans 25 États en 2014 (*Center for Educational Policy*, 2010). Pourtant, la question de rendre l'obtention du diplôme conditionnelle à la réussite au test, comme il est proposé dans certaines unités de sens, pose d'autres défis. Par exemple, il est possible, dans un tel contexte, que plusieurs élèves soient motivés plutôt par le risque d'avoir à reprendre l'épreuve, de redoubler ou de suivre des cours d'appoint que par la pratique de pondération (Flores & Clark, 2003 ; Stone, 2008). Aussi faudra-t-il considérer d'accompagner le test d'un projet sommatif mettant en vedette les habiletés supérieures de la pensée afin de compenser les limites des tests sommatifs à grande échelle sur le plan des habiletés et des compétences ciblées.

Selon les propos des répondants, la validité apparente du test peut constituer un troisième aspect de l'enjeu rattaché au test. Les élèves qui ne connaissent pas ou qui saisissent mal ce que mesure le test risquent ne pas le prendre au sérieux. De l'avis des répondants, plus le test prend du temps à le terminer et paraît long aux yeux des élèves, moins ils sont motivés à le remplir. Ceci serait surtout le cas chez les élèves à risque. Certains écrits estiment en effet que les élèves ont tendance à prendre davantage au sérieux une épreuve lorsqu'ils connaissent son but, son rôle, et lorsque celui-ci paraît approprié pour

ce qui est de la longueur, de la durée, de la familiarité du vocabulaire et de la difficulté des items (Nichols & Berliner, 2007). De plus, lorsque les élèves reçoivent de la rétroaction immédiate de leur enseignant, ils auraient également tendance à valoriser davantage le test (Luce-Kapler & Klinger, 2005).

Le deuxième facteur du modèle d'Eklöf (2010), c'est-à-dire le format des items, fait aussi l'objet des propos de répondants. L'OQRE laisse le choix de considérer seulement soit les questions à choix multiples, soit les questions à réponses élaborées ou soit une combinaison des deux, en partie ou dans leur ensemble, dans le calcul de la note finale au cours de mathématiques. Faute d'un corrigé ou d'un guide de notation, le personnel enseignant a tendance à choisir plutôt les questions à choix multiples en raison de leur facilité de correction. Toutefois, les membres sont conscients que ce choix risque de ne refléter qu'un échantillon étroit du curriculum et de ne pas tenir compte des styles cognitifs des élèves. Dans plusieurs situations, le choix du type d'items dépendrait également des conditions d'administration restreintes du test, un sujet traité dans les paragraphes suivants.

En effet quatre conditions d'administration, le troisième facteur du modèle d'Eklöf (2010), ressortent des commentaires des répondants : la communication de la pratique de pondération aux élèves, le moment de l'administration, la période d'accès au test et le niveau décisionnel. Le besoin de renseigner les élèves de manière transparente quant à la taille de la pondération et aux composantes notées aux fins de la note finale du cours semble poser un dilemme chez les répondants à cette étude. Les participants craignent que si l'élève sait que seules les questions à choix multiples compteront, il aura tendance à omettre les questions à développement, et s'il sait que seulement les questions à réponses élaborées comptent, il sera porté à deviner les réponses aux questions à choix multiples. Il est proposé comme solution de rechange de ne fournir à l'élève que des renseignements vagues au sujet des sections du test qui seront pondérées. Mais cette solution présente un véritable dilemme. Les répondants sont conscients qu'un examen incomplet ou qu'une sous-performance menacent la validité des résultats du test et ressentent le besoin d'encourager tous les élèves, en particulier ceux à risque, à remplir l'ensemble du test afin d'obtenir une note élevée dans leur cours de mathématiques. Ceci reflète une situation éthique difficile à réconcilier.

La deuxième condition d'administration qui revient dans les propos des répondants concerne le moment de l'administration du test. En Ontario, comme dans plusieurs autres instances juridiques d'éducation, les épreuves à

grande échelle sont administrées vers la fin de la session à la période où se donnent les épreuves sommatives des cours du palier secondaire. Cette situation présente deux difficultés. D'abord, un dédoublement d'examens sommatifs administrés dans une période relativement courte a déjà été évoqué plus haut. Ensuite les résultats des épreuves ne sont rendus publics qu'à la prochaine session ou la prochaine année du secondaire lorsque les élèves ont été promus à un niveau supérieur et ont changé d'enseignants. À ce stade, les résultats obtenus et toute rétroaction connexe perdent leur valeur. Afin de contrer ces difficultés, certains États américains administrent leurs tests plus tôt dans la session malgré le fait que les programmes d'études ne soient pas encore complètement achevés, ce qui constitue une autre problématique en soi (Luce-Kapler & Klinger, 2005).

Les unités de sens des répondants ont également fait ressortir l'accès limité au test, une troisième condition d'administration. En effet, elles se réfèrent au besoin de consulter de près au préalable le test afin de choisir les sections ou les items mêmes qui seront pondérés en fonction de ce qui a été étudié en classe, du vocabulaire utilisé, des exigences de correction et du niveau de difficulté. Dans sa thèse de doctorat de nature qualitative, Koch (2011) explique que les enseignants participant à son étude éliminaient les items qui présentaient un format, un contenu ou un vocabulaire différents de ce qui était vu en classe. Le besoin d'avoir au préalable un plus grand accès au test pour ce qui est du temps peut donner lieu à des notes élevées au bulletin, mais à quel prix lorsque seules les questions à choix multiples ou seules les questions de numération sont pondérées. Les questions reliées à cette problématique se multiplient dans le contexte de substitution de tests sommatifs aux fins du cours ou dans les conditions de réussite afin d'obtenir le diplôme d'études secondaires. Il s'agit ici d'établir un équilibre entre le besoin du personnel enseignant d'examiner le test au préalable, afin de prendre des décisions éclairées sur les items ou les sections qui serviront à la pondération, et le besoin de sécuriser le test et d'éliminer toute possibilité de fuite. D'où le besoin également de considérer, parallèlement, le niveau décisionnel. En effet, de l'avis des répondants, plus les enjeux sont élevés plus les décisions devraient se prendre à un échelon administratif élevé et centralisé.

Enfin, les propos des répondants portent sur plusieurs aspects reliés à l'enjeu du test, à son format et aux conditions d'administration mais ne se réfèrent aucunement à la charge cognitive des items, le dernier facteur du modèle d'Eklöf (2010) axé sur les caractéristiques du test. En effet, les commentaires

ne laissent pas entendre les difficultés pouvant être reliées au niveau de langue présent dans les items et à la présence de mots de vocabulaire non familiers ou non étudiés en classe, sauf dans le cas du besoin de choisir au préalable les items en fonction de ce qui a été vu durant le cours de mathématiques ou dans le cas des items qui font obstacle aux élèves à risque.

Conclusion

D'une population de 314 membres enseignants francophones qui ont participé aux sessions d'administration du test de mathématiques, 9^e année, de l'OQRE à l'hiver et au printemps 2010, 70% ont rempli la dernière page du questionnaire contextuel à leur intention où se trouvaient sept questions, dont une à composante fermée et à développement, au sujet de la pratique de pondération des résultats du test aux notes finales du cours de mathématiques. Selon les réponses à cette question, le personnel enseignant estime que cette pratique motive les élèves à remplir le test mais exprime plusieurs réserves à l'égard de la réaction des élèves face au test et les considérations liées au test et aux conditions de son administration.

Bien que les résultats de l'étude proviennent d'analyses secondaires des données d'une population enseignante particulière en réponse à une enquête et à un questionnement régis principalement par l'OQRE à des fins d'imputabilité plutôt que par une recherche fondamentale et qu'ils s'inspirent en majeure partie des commentaires à une seule question, les préoccupations qui en ressortent rejoignent les résultats des écrits et ceux inventoriés dans le modèle d'Eklöf (2010). Ils reflètent la nature multidimensionnelle et précaire de la pratique et le besoin d'examiner davantage les composantes et les implications sur les plans politique et académique. Il pourrait s'avérer intéressant, par exemple, d'étudier les répercussions d'une politique qui mettrait en œuvre la pratique hypothétique suivante : les plus hautes instances ministérielles recommandent que toutes les questions à choix multiples et qu'une portion des questions à réponse construite du test provincial représentent 15% de la note finale du cours de mathématiques, servent d'examen objectif de fin de session et soient accompagnées d'un projet mesurant les habiletés supérieures de la pensée et les compétences de communication. L'organisme responsable sélectionnerait au préalable les questions représentatives des domaines et des habiletés visées, fournirait les consignes de notation de ces questions et déterminerait les critères d'adaptation et d'exemption des élèves à risque. Il est facilement envisageable d'imaginer plusieurs autres scénarios

possibles. Par ailleurs, les commentaires fournissent à la communauté académique une richesse incontournable de pistes d'articulation d'hypothèses de recherche empirique qui serviront à établir et à valider un modèle de la motivation à remplir des évaluations à grande échelle en présence d'incitatifs non financiers à court et à long termes.

RÉFÉRENCES

- Barker, B. (2004). Teaching and targets: Self-evaluation and school improvement. *Educational Review*, 56, 210-211.
- Baumert, J., & Demmrich, A. (2001). Test motivation in the assessment of student skills: The effects of incentives on motivation and performance. *European Journal of Psychology of Education* 16(3), 441-462. doi: 10.1007/BF03173192
- Blau, J. R., Moller, S., & Jones, L. V. (2004). Why test? Talent loss and enrolment loss. *Social Science Research*, 33, 409-434. doi: 10.1016/j.ssresearch.2003.09.002
- Brown, R. (1999). Creating school accountability reports. *School Administrator*, 56(10), 12-14.
- Center for Educational Policy (2010). *State High School Tests: Exit exams and Other Assessments*. Retrieved from www.cep-dc.org/index.cfm?DocumentTopicID=7
- Coburn, C. E., & Talbert, J. E. (2006). Conceptions of evidence use in school districts: Mapping the terrain. *American Journal of Education*, 112(4), 469-495. doi: 10.1086/505056
- Copland, M. A. (2003). Leadership of inquiry: Building and sustaining capacity for school improvement. *Educational Evaluation and Policy Analysis*, 25, 375-395. doi: 10.3102/01623737025004375
- DeMars, C. (2000). Test stakes and item format interactions. *Applied Measurement in Education*, 13, 55-77. doi: 10.1207/s15324818ame1301_3
- Drasgow, F., Levine, M., & Williams, E. (1985). Appropriateness measurement with poly-chotomous item response models and standardized indexes. *British Journal of Mathematical & Statistical Psychology*, 38, 67-86.
- Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values and goals. *Annual Reviews*, 53, 109-132.
- Eklöf, H. (2010). Skill and will: test-taking motivation and assessment quality. *Assessment in Education: Principles, Policy & Practice*, 17(4), 345-356. doi: 10.1080/0969594X.2010.516569
- Ferrer-Caja, E., & Weiss, M. R. (2002). Cross-validation of a model of intrinsic motivation with students enrolled in high school elective courses. *Journal of Experimental Education*, 71, 41-65. doi: 10.1080/00220970209602056
- Flores, B. B., & Clark, E. R. (2003). Texas voices speak out about high-stakes testing: Preservice teachers, teachers, and students. *Current Issues in Education*, 6(3). Retrieved from <http://cie.ed.asu.edu/volume6/number3/>
- Glas, C. A. W., & Falcón, J. C. (2003). A comparison of item-fit statistics for the three-parameter logistic model. *Applied Psychological Measurement*, 27, 87-106. doi: 10.1177/014662160225053
- Glas, C. A. W., & Meijer, R. R. (2003). A Bayesian approach to person fit analysis in item response theory models. *Applied Psychological Measurement*, 27, 217-233. doi: 10.1177/0146621603027003003

- Gustafsson, J. E. (2008). Effects of international comparative studies on educational quality of educational research. *European Educational Research Journal*, 7, 1-17. doi: 10.2304/eeerj.2008.7.1.1
- Gustafsson, J. E., & Rosén, M. (2005). The dimensional structure of reading assessment tasks in the IEA Reading Literacy Study 1991 and the Progress in International Reading Literacy Study 2001. *Educational Research and Evaluation*, 12(5), 445-468. doi: 10.1080/13803610600697179
- Harlen, W., & Crick, R. D. (2003). Testing and motivation for learning. *Assessment in Education*, 10, 169-207. doi: 10.1080/0969594032000121270
- Hoyt, J. E. (2001). Performance funding in higher education: The effects of student motivation on the use of outcomes tests to measure institutional effectiveness. *Research in Higher Education*, 42, 71-85.
- Kerr, K. A., Marsh, J. A., Ikemoto, G. S., Darilek, H., & Barney, H. (2006). Strategies to promote data use for instructional improvement: Actions, outcomes, and lessons from three urban districts. *American Journal of Education*, 112(4), 496-520. doi: 10.1086/505057
- Klinger, D. A., DeLuca, C., & Miller, T. (2008). The evolving culture of large-scale assessments in Canadian education. *Canadian Journal of Educational Administration and Policy*, 76, 1-34.
- Koch, M. (2011). *Implications of the multiple-use of large-scale assessments for the process of validation: A case study of the multiple-use of a Grade 9 mathematics assessment* (Unpublished doctoral dissertation). University of Ottawa, Canada.
- Kornhaber, M. L. (2004). Appropriate and inappropriate forms of testing, assessment, and accountability. *Educational Policy*, 18(1), 45-70. doi: 10.1177/0895904803260024
- Lafontaine, D., & Simon, M. (2008). Évaluation des systèmes éducatifs. *Mesure et Évaluation en Éducation*, 31(3), 95-123.
- Levine, M., & Rubin, D. (1979). Measuring the appropriateness of multiple-choice test scores. *Journal of Educational Statistics*, 4(4), 269-290. doi: 10.2307/1164595
- Levitt, S. D., List, J. A., Neckermann, S., & Sadoff, S. (2011). *The Impact of Short-term Incentive on Student Performance*. Mimeo, University of Chicago, Becker Friedman Institute. Retrieved from http://bfi.uchicago.edu/events/20111028_experiments/papers/Levitt_List_Neckermann_Sadoff_Short-Term_Incentives_September2011.pdf
- Loye, N. (2011). Panorama des programmes actuels des enquêtes à large échelle. *Mesure et Évaluation en Éducation*, 34(2), 3-24.
- Luce-Kapler, R., & Klinger, D. (2005). Uneasy writing: The defining moments of high-stakes literacy testing. *Assessing Writing*, 10, 157-173. doi: 10.1016/j.asw.2005.08.002
- Ndinga, P., & Frenette, E. (2010). Élaboration et validation de l'Échelle de motivation à bien réussir un test (ÉMRT). *Mesure et Évaluation en Éducation*, 33(3), 99-123.
- Nering, M. (1997). The distribution of indexes of person fit within the computerized adaptive testing environment. *Applied Psychological Measurement*, 21, 115-127. doi: 10.1177/01466216970212002
- Nichols, S. L., & Berliner, D. C. (2007). *Collateral Damage: How High-Stakes Testing Corrupts America's Schools*. Cambridge, NE: Harvard Education Press.
- Olaru, I. D., Karageorghis, C. I., & Chatzisarantis, N. K. D. (2003). Multiple motivation theory testing in secondary school physical education. *Journal of Sports Sciences*, 21, 356-357.

- O'Neil, H. F., Abedi, J., Miyoshi, J., & Mastergeorge, A. (2005). Monetary incentives for low-stakes tests. *Educational Assessment, 10*(3), 185-208. doi: 10.1207/s15326977ea1003_3
- Radford, L., & Demers, S. (2010). Pratiques pédagogiques et rendement des élèves au Test Provincial des Mathématiques de l'Ontario. Document inédit, Centre GIARE, Ottawa, Canada.
- Roderick, M., & Engel, M. (2001). The grasshopper and the ant: Motivational responses of low-achieving students to high-stakes testing. *Educational Evaluation and Policy Analysis, 23*(3), 197-227. doi: 10.3102/01623737023003197
- Sloane, F. C., & Kelly, A. E. (2003). Issues in high-stakes testing programs. *Theory into Practice, 42*, 12-17. doi: 10.1353/tip.2003.0011
- Stocking, M., Steffen, M., & Eignor, D. (2001). *A method for building a realistic model of test taker behavior for computerized adaptive testing*. Princeton, NJ: Educational Testing Service.
- Stone, K. R. (2008). *Making sense of testing: English language learners and statewide assessment* (Unpublished Doctoral Dissertation). University of Minnesota, Minnesota.
- Van Barneveld, C., King, S., Simon, M., & Nadon, C. (2010). *Final Report on the Analysis of Teachers' Responses to Constructed-Response Questions on the Grade 9 Assessment of Mathematics Teacher Questionnaire. Winter and Spring 2010 administration*. Toronto, Canada: EQAO.
- Van Barneveld, C., Pharand, S.-L., Ruberto, L., & Haggarty, D. (2012). Student motivation in large-scale assessments. In M. Simon, K. Ercikan, & M. Rousseau (Eds). *Improving Large-Scale Assessment In Education* (p. 43-61). New York, NY: Routledge, Taylor & Francis Group.
- Van Krimpen-Stoop, E., & Meijer, R. (1999). The null distribution of person-fit statistics for conventional and adaptive tests. *Applied Psychological Measurement, 23*, 327-345. doi: 10.1177/01466219922031446
- Volante, L. (2008). Educational Quality and Accountability in Ontario: Past, present and future. *Canadian Journal of Educational Administration and Policy Issue, 58*, 1-21.
- Wayman, J. C., & Stringfield, S. (2006a). Data use for school improvement: School practices and research perspectives. *American Journal of Education, 112*(4), 463-468. doi: 10.1086/505055
- Wayman, J. C., & Stringfield, S. (2006b). Technology-supported involvement of entire faculties in examination of student data for instructional improvement. *American Journal of Education, 112*(4), 549-571. doi: 10.1086/505059
- Wigfield, A., & Eccles, J. (2000). Expectancy-value theory of achievement motivation. *Contemporary Educational Psychology, 25*, 68-81. doi: 10.1006/ceps.1999.1015
- Wise, L. L. (1996a, April). *A persistence model of motivation and test performance*. Paper presented at the annual meeting of the American Educational Research Association, New York, NY.
- Wise, L. L. (1996b). *Indicators of student effort on the National Assessment of Educational Progress* (Report No. R99B60002). Alexandria, VA: Human Resources Research Organization.
- Wise, S. L. (2006). An investigation of the differential effort received by items on a low-stakes computer-based test. *Applied Measurement in Education, 19*(2), 95-114.

- Wise, S. L., & DeMars, C. E. (2005). Examinee motivation in low-stakes assessment: Problems and potential solutions. *Educational Assessment, 10*, 1–17.
- Wolf, L. F., & Smith, J. F. (1995). The consequence of consequence – motivation, anxiety, and test- performance. *Applied Measurement in Education, 8*, 227-242. doi: 10.1207/s1532418ame0803_3
- Wolf, L. F., Smith, J. F., & Birnbaum, M. E. (1995). Consequence of performance, test motivation, and mentally taxing items. *Applied Measurement in Education, 8*, 341-351. doi: 10.1207/s153241818ame0804_4
- Young, V. M. (2006). Teachers' use of data: Loose coupling, agenda setting, and team norms. *American Journal of Education, 112*(4), 521-548. doi : 10.1086/505058

Date de réception : 27 février 2012

Date de réception de la version finale : 18 décembre 2012

Date d'acceptation : 20 décembre 2012