

*COMMENTAIRE DE LA
RÉDACTRICE INVITÉE*

**PROGRAMMES
D'ÉVALUATION DU
RENDEMENT EN LECTURE
ET ÉCRITURE : QUELLES
LEÇONS EN TIRER?**

Marielle Simon
Université d'Ottawa
Ottawa, Ontario

Le Canada et ses instances provinciales et territoriales souscrivent régulièrement à des programmes d'évaluation du rendement des élèves, principalement à des fins d'imputabilité sociale. En plus d'administrer leur propres programmes provinciaux ou territoriaux, depuis 30 ans, ils participent, entre autres, à la Troisième enquête internationale sur l'enseignement des mathématiques et des sciences (TEIEMS), au Programme international de la recherche en lecture scolaire (PIRLS), au Programme international pour le suivi des acquis (PISA), et au Programme pancanadien d'indicateurs du rendement scolaire (PIRS), remplacé récemment par le Programme pancanadien d'évaluation (PPCE). Ces programmes évaluent le rendement des élèves du primaire et du secondaire en mathématiques, en sciences, en lecture, et en écriture. Le PIRLS, le PISA, et le PPCE ont tous le volet lecture, mais seul le PIRS offre la composante écriture.

*GUEST EDITOR'S
COMMENT*

**ACHIEVEMENT ASSESS-
MENT PROGRAMS IN
READING AND WRITING:
WHAT LESSONS CAN BE
LEARNED?**

Marielle Simon
University of Ottawa
Ottawa, Ontario

Canada and its provinces and territories regularly subscribe to student achievement programs, primarily for social accountability purposes. In addition to administering their own provincial and territorial programs, they have participated for the past 30 years in the Third International Mathematics and Science Study (TIMSS), Progress in International Reading Literacy Study (PIRLS), Programme for International Student Assessment (PISA), and School Achievement Indicators Program (SAIP), recently replaced by the Pan-Canadian Assessment Program (PCAP). These programs assess the achievement of primary and secondary students in mathematics, science, reading, and writing. The PIRLS, PISA and PCAP all have a reading component, but only SAIP has a section on writing.

Correspondance à l'auteure : Marielle Simon, Faculté d'Éducation, Université d'Ottawa, 145 Jean-Jacques Lussier, CP 450 Succursale A, Ottawa, ON, K1N 6N5; <msimon@uottawa.ca>

Ces programmes donnent d'abord lieu à de nombreuses comparaisons des résultats entre instances, entre groupes linguistiques, et entre cycles d'application ainsi qu'en termes de seuils de réussite. Au Canada, la comparaison des résultats et entre groupes linguistiques s'avère une pratique courante surtout en raison de l'administration de ces programmes dans les deux langues officielles du pays. Les médias, les responsables des programmes d'évaluation, et certains chercheurs rapportent des écarts importants entre le rendement des élèves des écoles de langue française en milieu minoritaire et ceux des écoles anglaises en milieu majoritaire (Bussière et al., 2001; Bussière, Cartwright, Knighton, & Rogers, 2004; Landry & Allard, 2002). Toutefois, la justesse de toute comparaison dépend du respect des principes d'équité préconisés dans le domaine de la mesure et de l'évaluation. Les documents *Standards for Educational and Psychological Testing* (American Educational Research Association, Association for Psychological Assessment, & National Council on Measurement in Education, 1999), *Student Evaluation Standards* (Joint Committee on Standards for Educational Evaluation, 2003), et *Principes d'équité relatifs aux pratiques d'évaluation des apprentissages scolaires au Canada* (Comité consultatif mixte, 1993) ont pour but de guider l'élaboration de programmes d'évaluation de qualité. Ils prescrivent des règles strictes concernant les besoins de comparaisons entre populations.

Les programmes d'évaluation du rendement s'avèrent également une

First and foremost, these programs give rise to many comparisons of results between jurisdictions, linguistic groups, years of administration, and expected levels. In Canada, comparing results between language groups, particularly students from English and French schools, is a common practice because the programs are administered in both of the country's official languages. Media, assessment program administrators, and some researchers have reported major variances in student achievement in French-language schools in minority communities and English schools in a majority situation (Bussière et al., 2001; Bussière, Cartwright, Knighton, & Rogers, 2004; Landry & Allard, 2002). However, the soundness of any comparison depends on following the equity principles recommended for measurements and assessments. The purpose of the *Standards for Educational and Psychological Testing* (American Educational Research Association, Association for Psychological Assessment, & National Council on Measurement in Education, 1999), *Student Evaluation Standards* (Joint Committee on Standards for Educational Evaluation, 2003), and *Principles for Fair Student Assessment Practices for Education in Canada* (Joint Advisory Committee, 1993) is to guide the development of high-quality assessment programs. They prescribe strict rules for comparisons between populations.

source inépuisable de données pouvant alimenter les décisions administratives, politiques, financières, et sociales. En effet, plusieurs pays effectuent des analyses avancées des résultats de ces programmes en vue de mieux comprendre les déterminants du rendement scolaire et de proposer ainsi des plans de redressement appropriés. Malgré des investissements considérables, ces enquêtes s'avèrent peu utiles aux yeux des praticiens en milieu scolaire, des dirigeants en éducation, et des chercheurs universitaires, faute, en grande partie, de la qualité des résultats. La qualité dépend dans une grande mesure du niveau d'engagement des élèves et de leurs enseignants au moment de l'administration des épreuves et des questionnaires contextuels. Ceci est davantage vrai plus les enjeux pour l'élève, ou pour l'école, sont importants. Cette problématique suscite une foule de questions. « Dans quelle mesure les élèves et leurs enseignants appelés à participer à ces programmes les trouvent-ils utiles et signifiants? » « Comment perçoivent-ils leur rôle et leur influence sur la qualité des données qu'ils fournissent? » « Compréhendent-ils les retombées de leur participation à ces programmes? » « Comment peut-on les engager davantage à fournir un rendement optimal à ces programmes? ».

Une autre série de questions concernent les dimensions de la langue évaluées. « Dans quelle mesure les savoirs théoriques et les pratiques au plan de l'apprentissage et de l'enseignement de la lecture et de l'écriture se rapprochent-elles de ceux adoptés par les programmes d'évaluation? » « Dans

Student achievement assessment programs also prove an inexhaustible source of data on which to base administrative, political, financial, and social decisions. In fact, many countries conduct advanced analyses of the results of these programs to better understand the determinants of school achievement and propose appropriate adjustment plans. Despite substantial investments, the assessments are not very useful to school practitioners, educational administrators, and university researchers, mainly because of the quality of the results. Their quality depends primarily on the level of commitment of students and their teachers at the time the tests and background questionnaires are administered. This is even more true when the stakes for students or schools are high. This issue raises a number of questions. "To what extent do the students and teachers who take part in these programs find them useful and meaningful?" "How do they perceive their role and influence on the quality of the data they provide?" "Do they understand the benefits of participating in these programs?" "How can students and teachers be further encouraged to perform optimally in these programs?"

Another series of questions focuses on the definition of the language skills assessed. "To what extent are theoretical knowledge and practices in learning and teaching reading and writing aligned with the knowledge and practices adopted

quelle mesure les concepteurs des programmes d'évaluation, les politiciens, les chercheurs, et les praticiens s'entendent-ils dans leur définition de l'enseignement, l'apprentissage, et l'évaluation de la lecture et de l'écriture? ». Ce questionnement au sujet des concepts de la lecture et d'écriture ainsi que l'utilité des programmes d'évaluation aux yeux des participants, des dirigeants, et des chercheurs revient à l'étude de leur validité.

La motivation sous-jacente à l'organisation du numéro spécial de la Revue s'inspire d'un atelier bilingue international qui a eu lieu en juin 2006 et qui portait sur l'utilité des programmes d'évaluation du rendement en lecture et en écriture. L'atelier a réuni une vingtaine de participants invités provenant des milieux scolaires, ministériels, et universitaires en vue d'échanger et de recommander des solutions de rechange qui pourraient rendre les programmes d'évaluation du rendement davantage utiles, attrayants, et valides aux yeux de tous les partis concernés. Il proposait de répondre aux trois questions suivantes. (a) Les définitions de la compétence en lecture et en écriture selon les intervenants dans les écoles, les concepteurs des programmes d'évaluation à grande échelle, les chercheurs dans le domaine de la lecture et de l'écriture, et les utilisateurs des données et les résultats de ces évaluations sont-elles congruentes entre elles? (b) Dans quelle mesure les programmes d'évaluation de la lecture et de l'écriture tiennent-ils compte de ces définitions? (c) Comment peut-on rendre ces programmes d'évaluation davantage attrayants, accessibles,

by assessment programs?" "To what extent do assessment program designers, politicians, researchers, and practitioners agree on their respective definitions of teaching, learning, and assessment of reading and writing?" Questioning the underlying concepts of reading and writing, as well as the usefulness of assessment programs from the perspective of participants, administrators, and researchers essentially involves examining their validity.

The motivation for the Journal's special issue came from a bilingual international workshop held in June 2006, which focused on the usefulness of assessment programs in reading and writing achievement. The workshop brought together some 20 invited participants from the school, government, and university sectors to discuss and recommend alternative solutions that could make achievement assessment programs more useful, attractive, and valid from the perspective of the various stakeholders. The workshop aimed to answer the following three questions: (a) Are the definitions of reading and writing skills adopted by school practitioners, large-scale assessment program designers, researchers in reading and writing, and users of the assessment data and results congruent with one another?; (b) To what extent do reading and writing assessment programs take these definitions into account?; and (c) How can assessment programs be made more attractive, accessible, and useful for stake-

et utiles aux intervenants du milieu scolaire? L'atelier, d'une durée d'une journée et demie, s'articulait autour de ces thèmes sous forme de symposiums et de tables rondes.

Les résultats de l'atelier tels que rapportés dans ce numéro spécial devraient intéresser les concepteurs des programmes, les politiciens, et les dirigeants des unités scolaires des provinces canadiennes, du Canada, et de tout pays concerné. De surcroît, les résultats de l'atelier devraient également nourrir les initiatives voulant que les pratiques d'évaluation rejoignent les fondements du nouveau paradigme en éducation basé sur la perspective socio-constructiviste et des théories sous-jacentes à l'apprentissage de la lecture et de l'écriture. Aussi devraient-ils permettre de solidifier le concept de validité des programmes d'évaluation. Les réponses aux questions posées lors de l'atelier visent l'arrimage entre la théorie et la pratique de l'enseignement et de l'évaluation de la lecture et de l'écriture. Le commentaire qui suit fait ressortir les leçons à tirer du numéro spécial au plan des conceptions de la lecture et de l'écriture, de la nature et de la qualité de la mesure, et de l'utilité des programmes d'évaluation aux fins des politiques et de la pratique.

LEÇONS À TIRER AU PLAN DES CONCEPTIONS DE LA LECTURE ET DE L'ÉCRITURE

Les articles de Hansen et de Klinger et Luce-Kapler offrent des leçons à tirer par rapport aux conceptions de la lecture et de l'écriture du point de vue

holders in school communities? At the one-and-a-half-day workshop, the various topics were examined in symposiums and round-table discussions.

The results of the workshop reported in this special issue should be of interest to assessment program designers, politicians, and school administrators in Canadian provinces, Canada as a whole, and any other country involved in large-scale assessment. Moreover, the workshop results should encourage initiatives to align assessment practices with the new educational paradigm based on the socio-constructivist approach and the underlying theories of learning to read and to write and also strengthen the construct validity of assessment programs. The answers to the questions raised at the workshop are intended to align theory and practice in teaching and assessment of reading and writing. The following comments identify lessons that can be learned from this special issue on reading and writing concepts, the nature and quality of the assessment, and the usefulness of assessment programs for policy and practice.

LESSONS LEARNED ON READING AND WRITING CONCEPTS

Lessons can be learned from the articles by Hansen and Klinger and Luce-Kapler on the conception of reading and writing from the perspective of teachers, researchers

des enseignants, des chercheurs dans la discipline, et des élèves. Hansen se penche plus particulièrement sur les valeurs et les croyances qu'ont les enseignants à l'endroit de l'apprentissage de la lecture et de l'écriture et se demande dans quelle mesure l'évaluation en salle de classe et celle à grande échelle tiennent-elles compte de ces valeurs. Elle se positionne en faveur d'une pratique d'évaluation qui appuie l'apprentissage plutôt que d'en déterminer son bilan. Bien que les conditions de la salle de classe et celles des programmes d'évaluation à grande échelle diffèrent à plusieurs plans, l'article de Hansen permet de tirer les leçons suivantes : offrir des choix à l'élève afin de l'engager, de rejoindre ses intérêts, et de l'amener à s'investir et à développer sa propre voix comme écrivain et comme lecteur; valoriser l'unicité de l'élève dans ses écrits et dans ses lectures; et placer l'élève dans des situations d'écriture ou de lecture qui lui permettent de continuer à apprendre et à se développer. Elle préconise la notion de culture d'apprentissage de la lecture et de l'écriture où les élèves s'entraident et font appel à des stratégies métacognitives comme l'autoévaluation. Elle s'éloignerait de la notion du nivellement de la lecture en faveur d'une plus grande spontanéité semblable au contexte de l'apprentissage de l'écriture.

Par ailleurs, Klinger et Luce-Kapler rapportent une étude empirique menée auprès d'élèves de 10^e année en Ontario qui doivent réussir le Test provincial de compétences linguistiques (TPCL) afin d'obtenir leur diplôme d'études secondaires. Ils s'intéressent

in the field, and students. Hansen specifically focuses on the values and beliefs teachers have about learning to read and write and asks to what extent large-scale and classroom assessments take these values into account. She favours an assessment practice *for* learning rather than *of* learning. Although the conditions of classroom and large-scale assessment programs differ in many ways, Hansen's article teaches the following lessons: offer choices that engage students, match their interests, and encourage self-investment and the development of their own voices as readers and writers; value students' unique style in writing and reading; and provide students with reading and writing situations that enable them to continue to learn and develop. Hansen promotes a learning culture for reading and writing where students help each other and use metacognitive strategies such as self-assessment. She would move away from the reading level concept in favour of greater spontaneity, similar to the situation for learning to write.

In their article, Klinger and Luce-Kapler report on an empirical study conducted among Grade 10 students in Ontario who are required to pass the Ontario Secondary School Literacy Test (OSSLT) in order to receive their high-school diploma. The authors are interested in the consequential validity of high-stakes assessment programs and thus examine how students prepare for the test, its impact on

à la validité et à la conséquence des programmes d'évaluation à enjeux importants et examinent ainsi la nature de la préparation au test, l'impact de celui-ci sur les aspirations professionnelles des élèves, et la valeur que les élèves accordent au test. Ils comparent, à l'aide d'entrevues et de groupes de discussion, les perceptions des élèves qui ont réussi l'épreuve à celles qui l'ont échoué. Les élèves des deux groupes (réussite et non-réussite) ont tendance à trouver redondante et peu pertinente la préparation à l'épreuve et voient peu la pertinence de cette préparation lorsqu'elle se fait à l'extérieur du cours de langue. En ce sens, l'étude recommande d'examiner le besoin d'une préparation en vue de développer des stratégies de passation d'une épreuve à grande échelle peu importe sa discipline et d'insister sur une conception plus large des compétences linguistiques plutôt qu'uniquement d'écriture et de lecture. Les deux groupes exprimeraient des aspirations professionnelles différentes. Enfin, les élèves désirent obtenir de la rétroaction aux fins d'amélioration. Les auteurs soulèvent donc le besoin de communiquer clairement les buts et objectifs des programmes d'évaluation du rendement.

Dans son évaluation du livre de l'International Association for the Evaluation of Educational Achievement portant sur le cadre conceptuel du PIRLS, Peterson ancre sa position par rapport aux approches idéologique et individuelle (autonomie) du développement de la littératie. Selon elle, le livre présente des propos qui s'accordent tant vers l'une ou l'autre des deux approches. Elle mentionne toutefois, sans

students' occupational aspirations, and the value students attribute to the test. The perceptions of students who passed the test are compared to the perceptions of students who failed, using interviews and discussion groups. Students in both groups (pass and fail) tend to find test preparation useless and irrelevant and do not see the purpose of test preparation outside language arts classes. The study recommends examining the need for preparation to develop strategies for taking large-scale tests in any discipline and emphasizes a broader conception of language skills rather than just reading and writing. Lastly, although the two groups expressed different occupational aspirations, they all want feedback for improvement. The authors point out the need to clearly communicate the goals and objectives of achievement assessment programs.

In her review of the book from the International Association for the Evaluation of Educational Achievement on the conceptual framework of the PIRLS, Peterson describes its position on the ideological and autonomous approaches to literacy development. According to Peterson, the book provides information in support of both approaches. However, without going into detail, she mentions the lack of a sociocultural approach to literacy. Although reading and writing achievement assessment programs involve literacy, they must also take into account the theoretical perspective that underlies teaching and learn-

élaborer, l'absence de l'approche socioculturelle envers la littératie. Bien que les programmes d'évaluation du rendement en lecture et en écriture s'inscrivent dans un contexte de littératie, ils doivent également tenir compte de la perspective théorique qui sous-tend l'enseignement et l'apprentissage en général. Les modèles de mesure ont tendance à privilégier l'approche cognitive de l'apprentissage plutôt que socioconstructiviste qui fait davantage appel aux habiletés métacognitives et aux interactions sociales. En ce sens les programmes d'évaluation s'arriment peu avec les modèles d'enseignement et d'apprentissage préconisés dans les programmes d'études actuels. Ceci peut expliquer, dans une certaine mesure, la réaction face à la lecture de l'article de Hansen dans un contexte de programmes d'évaluation à grande échelle du rendement en lecture et en écriture.

LEÇONS À TIRER AU PLAN DE LA MESURE DE LA LECTURE ET DE L'ÉCRITURE

Les programmes d'évaluation ciblent d'abord des indicateurs qui sont par la suite opérationnalisés en mesures concrètes de la performance. Ceci se traduit habituellement en épreuves subies par les élèves dans les disciplines retenues et en questionnaires contextuels à l'intention des élèves, des enseignants, des directions d'écoles, et des parents. Cette opérationnalisation est une étape déterminante dans l'établissement de la validité des programmes. Elle doit mener à une collecte de données qui permette de se prononcer

ing in general. Measurement models tend to favour a cognitive rather than a socioconstructivist approach to learning, which emphasizes metacognitive skills and social interaction. In that sense, assessment programs are not aligned with the teaching and learning models recommended in current curriculum programs. This may explain, to a certain extent, the reaction readers may have to Hansen's article in a context of large-scale assessment programs of achievement in reading and writing.

LESSONS LEARNED FROM READING AND WRITING MEASURES

Assessment programs first identify indicators that then become concrete measures of achievement, generally tests taken by students in selected subjects and background questionnaires for students, teachers, school administrators, and parents. How the programs operationalize skills is key to establishing program validity and should result in the collection of data that produce determinative findings on the components of the design framework and the selected development standards.

In their article on the development of assessment programs to inform policy and practice, Ericikan and Barclay-McKeown discuss five criteria for developing assessment programs for practical purposes: importance of skills assessed, align-

sur les éléments du cadre conceptuel et des normes d'élaboration retenus.

Dans leur article sur l'élaboration de programmes d'évaluation dans le but d'éclairer les politiques et la pratique, Ercikan et Barclay-McKeown discutent de cinq critères d'élaboration de programmes d'évaluation à des fins pratiques : mesure de compétences valorisées, arrimage de cette mesure avec les buts pédagogiques retenus, exactitude de la mesure, nature des rapports d'évaluation, et anticipation des répercussions. Les auteures signalent des limites importantes associées aux programmes d'évaluation mais elles approfondissent peu leurs propos. Bien que les buts pédagogiques retenus dans les programmes d'études et dans les programmes d'évaluation préconisent l'évaluation d'habiletés complexes en lecture et en écriture, les programmes d'évaluation offrent peu de modalités de questionnement. Dans les épreuves de lecture de l'Office de la qualité et de la responsabilité en éducation (OQRE) en Ontario, administrées en troisième et en sixième année, par exemple, les questions à choix multiples constituent jusqu'à 70 % de l'ensemble des items. Le reste comprend des questions ouvertes à réponses plutôt courtes. Le PIRLS présente en proportion égale les items à choix multiples et à réponses élaborées. Très peu de programmes d'évaluation ont recours à la manipulation de matériel concret comme fut le cas dans les années 1980 et 1990.

Selon Ercikan et Barclay-McKeown, les programmes d'évaluation doivent s'aligner avec les pratiques de la salle de

ment of assessment with selected instructional goals, measurement accuracy, nature of assessment reports, and anticipation of consequences. The authors point out major limitations associated with assessment programs without providing much detail. Although the instructional goals selected in curriculum and assessment programs advocate the assessment of complex reading and writing skills, assessment programs use few item types. In reading tests from the Education Quality and Accountability Office (EQAO) in Ontario, administered in Grade 3 and 6, for example, multiple-choice questions form up to 70% of all items. The rest are fairly short-answer, open-ended questions. The PIRLS has an equal proportion of multiple-choice and constructed response items. Very few assessment programs use manipulation of concrete material as was done in the 1980s and 1990s.

According to Ercikan and Barclay-McKeown, assessment programs should be linked to classroom practices, not the other way around, because teaching practices are generally based on learning and skill-development models, which are themselves based on empirical and theoretical research. In this regard, Ercikan and Barclay-McKeown rightly propose greater cooperation between learning experts, subject-matter experts, and assessment experts in order to develop more varied and relevant achievement measures.

classe et non l'inverse, puisque les pratiques d'enseignement s'appuient sur des modèles d'apprentissage et de développement de compétences fondés sur la recherche empirique et théorique. À cet effet, Ercikan et Barclay-McKeown ont raison de proposer une plus grande collaboration entre les experts en matière d'apprentissage, les experts en matière de la discipline ciblée, et les experts de la mesure en vue de développer des mesures de performance davantage variées et pertinentes.

En ce qui concerne l'exactitude de la mesure, Ercikan et Barclay-McKeown discutent de l'effet du nombre de niveaux de rendement dans l'échelle de notation sur l'exactitude du classement d'un élève étant donné le niveau de fidélité de l'épreuve. Selon elles, plus l'échelle possède de catégories, moins le classement est exact, et cette relation se complique en fonction de la fidélité de la mesure. Il va sans dire que la fidélité de toute épreuve dépend essentiellement du nombre et de la qualité des items qui la constituent en vue d'évaluer un certain nombre de compétences visées. Bref, l'exactitude de la mesure revient au besoin d'énoncer clairement les intentions de départ des programmes d'évaluation et les compétences ciblées afin de déterminer un nombre raisonnable d'items et de niveaux de rendement.

Ercikan et Barclay-McKeown passent ensuite à l'importance de rapporter des résultats utiles, significatifs, et opportuns. Cette recommandation s'applique en particulier aux programmes d'évaluation administrés à l'ensemble d'une population et qui accordent un ou

With respect to measurement accuracy, Ercikan and Barclay-McKeown discuss the effect of the number of achievement levels in the scoring scale on the accuracy of student ranking, based on the reliability of the test. According to the authors, the more categories a scoring scale has, the less accurate the ranking; and the relationship is further complicated by test reliability. It goes without saying that the reliability of any test depends essentially on the number and quality of the items that make up the test and the number of target skills. In short, measurement accuracy comes down to the need to clearly state the initial goals of assessment programs and the target skills in order to determine a reasonable number of items and achievement levels.

Ercikan and Barclay-McKeown then look at the importance of reporting useful, significant, and timely results. This recommendation applies, in particular, to assessment programs administered to an entire population that assign one or more score(s) to each student. According to the authors, programs where the results are in line with the grades obtained in the classroom are more relevant to stakeholders. They also underscore the importance of reporting not only the student's achievement results, but a variety of indicators, such as the sociocultural level of the school.

In discussing their experience with secondary analyses of the data from

plusieurs scores à chaque élève. Selon elles, les programmes dont les résultats s'accordent avec les notes obtenus en salle de classe s'avèreraient davantage pertinents aux yeux des intervenants. À cette fin, elles soulignent également l'importance de rapporter non seulement des résultats de la performance de l'élève mais une diversité d'indicateurs comme le niveau socioculturel de l'école.

En relatant leur expérience lors de leurs analyses secondaires des données de divers programmes d'évaluation en vue de déterminer les variables explicatives du rendement de l'élève, Simon, Roberts, Tierney, et Forgette-Giroux relèvent des défis pratiques, conceptuels, techniques, et statistiques. Au plan de la mesure, l'écart entre les variables d'intérêt aux chercheurs aux fins d'analyses secondaires et la nature des questions des programmes d'études constitue un défi de taille. S'ajoutent à cela la nature et la pertinence des échelles de questionnement et de notation et les techniques d'échantillonnage. Ces défis se multiplient lors de l'exploration de données issues du contexte linguistique minoritaire au Canada. Les chercheurs en général ont donc intérêt à proposer des modèles statistiques plus simples et qui tiennent compte des limites des programmes d'évaluation afin de pouvoir tirer davantage profit des résultats à des fins pratiques et politiques.

Enfin, la note de Blais sur la recherche et les méthodes se penche plus particulièrement sur la problématique d'échantillon. Il note l'incidence du processus d'échantillonnage sur la va-

rious assessment programs to identify variables that explain student achievement, Simon, Roberts, Tierney, and Forgette-Giroux report various practical, conceptual, technical, and statistical challenges. In an assessment, the variance between the variables of interest to researchers conducting secondary analyses and the nature of the questions in curriculum programs poses a significant challenge, as do the nature and relevance of item and scoring scales as well as sampling techniques. Looking at data from minority language communities in Canada further adds to the challenge. Researchers are generally well-advised to propose simpler statistical models that take the limitations of assessment programs into account to use the results more effectively for practical and political purposes.

Lastly, the research and practice note by Blais specifically discusses sampling issues. He notes the impact of the sampling process on the variability of the students' average age and the participation rate from one population or educational jurisdiction to another. Blais urges caution in the use of achievement rankings of the various participating countries based on their respective results in various assessment programs.

riabilité de l'âge moyen des élèves et du taux de participation d'une population ou d'une instance juridique à l'autre. Ses propos suscitent la prudence dans la lecture des palmarès des pays selon leurs résultats respectifs aux divers programmes d'évaluation.

LEÇONS À TIRER SUR L'UTILITÉ DES RÉSULTATS DES PROGRAMMES D'ÉVALUATION

Dans leur article, Savard, Sévigny, et Beaudoin examinent les liens entre le score globale et les scores analytiques en écriture, et ce à partir des résultats des populations des deux langues officielles au Canada. Ils se demandent dans quelle mesure l'ensemble des scores aux diverses composantes de l'écriture réussit à prédire le score global. Leur étude puise ses données à partir du PIRS en écriture de 1994. Les auteurs constatent une relation significative entre le score global et l'ensemble des scores analytiques mais remarquent également que les variables *structures et règles* expliquent davantage le score global chez les élèves anglophones tandis que les variables *règles et vocabulaire* figurent au haut de la liste chez les élèves francophones. Leur étude fait preuve, dans une certaine mesure, de la validité de construit du score global défini en termes de situation, contenu, organisation, vocabulaire, structure, et règles.

De leur côté, Bouchamma, Lapointe, et Richard ont également appliqué des analyses secondaires sur les données du PIRS en écriture de 2002 afin d'identifier les variables scolaires déterminantes du rendement des

LESSONS LEARNED ABOUT THE USEFULNESS OF ASSESSMENT PROGRAM RESULTS

In their article, Savard, Sévigny, and Beaudoin examine the relationship between overall scores and analytical scores in writing based on the results of student subpopulations from both official languages in Canada. They question how much all the scores in the various writing components can predict the overall score. Their study uses the data from the 1994 SAIP assessment of writing. The authors observe a significant relationship between the overall score and all analytical scores but also notice that *structure and rules* variables better explain the overall scores among Anglophone students, whereas *rules and vocabulary* variables top the list in predicting Francophone students' overall scores. The study shows, to some extent, the construct validity of the overall score determined by the scores of writing assessment components such as tasks, content, organization, vocabulary, structure, and rules.

Bouchamma, Lapointe, and Richard also apply secondary analyses to 2002 SAIP writing data to identify school variables that determine student achievement in French-language schools in minority situations in Canada as well as student achievement in French-language schools in a majority situation in Quebec. The authors found that the school, particularly its human and

élèves des écoles de langue française en contextes minoritaires au Canada et du rendement des élèves des écoles de langue française du Québec, contexte majoritaire. Les auteurs rapportent que l'école, en particulier ses ressources humaines et matérielles, ses relations école-communauté, et la vision de sa direction, a une certaine incidence sur le rendement de l'élève en écriture. Ce lien se ferait davantage sentir dans les milieux linguistiques minoritaires. L'article toutefois renseigne peu sur les répercussions des résultats sur la gestion des écoles en milieux minoritaires ni sur les limites des programmes d'évaluation au plan des cadres conceptuels sous-jacents à la mesure de ce type de variables. Dans le cas du PIRS en écriture, le modèle de Wang, Haertel, et Walberg (1993) s'avère trop général, ce qui a pour effet de cerner difficilement les variables prédictives du rendement de l'élève dans cette matière. Cette constatation appuie le besoin de définir en collaboration les cadres conceptuels des épreuves et des questionnaires contextuels afin de répondre tant aux besoins des programmes d'évaluation que de ceux des analyses secondaires.

Dans son évaluation du livre rédigé par Bouchamma et Long, IsaBelle reconnaît la valeur pratique et politique des résultats d'analyses secondaires menées sur les données du PIRS en écriture. Effectivement, selon elle, les auteures ont su transformer les résultats d'analyses secondaires en pratiques pédagogiques et de supervision efficaces et utiles en salle de classe. Bien que ce livre serve d'exemple d'utilisation des résultats de programmes

material resources, school-community relationships, and the administration's vision, has some effect on student achievement in writing. The relationship appears to be more identifiable in minority-language situations. However, the article provides little detail on the impact of results on the administration of schools in minority communities or the limitations of assessment programs in terms of the underlying conceptual framework for measuring this type of variable. In the case of the SAIP writing assessment, the Wang, Haertel, and Walberg (1993) conceptual model proves to be too general, making it difficult to identify variables predictive of student achievement in this subject. That observation supports the need to collaboratively define the conceptual framework of tests and context-based questionnaires to meet the requirements of both assessment programs and secondary analyses.

In her evaluation of the book by Bouchamma and Long, IsaBelle acknowledges the practical and political value of the results of secondary analyses conducted using SAIP writing data. In fact, in her view, the authors were able to translate the results of secondary analyses into effective and useful instructional and supervisory practices for the classroom. Although the book serves as an example of the use of assessment program results for practical purposes, the validity of these findings still depends largely on measurement quality and appropriate and accurate analyses.

d'évaluation à des fins pratiques, il demeure que la validité de ces conclusions repose en grande partie sur la qualité des mesures et sur des analyses appropriées et exactes.

CONCLUSION

Ce numéro spécial rapporte plusieurs problématiques entourant les programmes d'évaluation de la lecture et de l'écriture et les leçons à en tirer. Il fait état des définitions des concepts de littératie, de lecture, et d'écriture qu'entretiennent les praticiens, les chercheurs, et les dirigeants et de l'écart entre les approches théoriques d'enseignement et d'apprentissage, des pratiques en salle de classe, et des programmes d'évaluation du rendement à grande échelle en matière de la lecture et de l'écriture. Les programmes d'évaluation ont tendance à réduire les compétences linguistiques à la compréhension et à la production de l'écrit dans des situations normalisées avec peu de rétroaction signifiante aux yeux des participants. Ceci a pour effet de limiter la motivation et l'engagement de l'élève à fournir son rendement maximal, à faire appel à ses habilités métacognitives, et à vouloir lire et écrire pour apprendre.

Au plan de la mesure, les programmes d'évaluation tiennent plus ou moins compte des théories de l'apprentissage en général et de la lecture et de l'écriture plus spécifiquement. Ils se basent plutôt sur certaines des attentes ciblées par les programmes d'études ou par la société, en particulier celles qui se prêtent aux questions à choix multiples ou aux questions à réponses courtes. Au

CONCLUSION

This special issue reports on many challenges surrounding reading and writing assessment programs and the lessons to be learned from them. It provides definitions of literacy, reading, and writing concepts used by practitioners, researchers, and administrators and identifies the gap between theoretical approaches to teaching and learning, teaching and learning practices in the classroom, and large-scale achievement assessment programs in reading and writing. Assessment programs tend to reduce language skills to understanding and producing written texts in standardized situations, with little meaningful feedback from the participant's viewpoint, which limits student motivation and commitment to maximal achievement, use of metacognitive skills, and desire to learn to read and write in order to learn.

Assessment programs more or less take learning theory in general, and reading and writing theory specifically, into consideration. Rather, they are based on target requirements of the curriculum program or on societal expectations, in particular those that lend themselves to multiple-choice and short-answer questions. Instead of drawing on socioconstructivist learning practices, assessment programs tend to encourage teachers to adopt the underlying cognitive models (Simon, Turcotte, & Forgette-Giroux, 2006). Measurement limitations,

lieu de s'inspirer des pratiques socio-constructivistes de l'apprentissage, les programmes d'évaluation ont tendance à amener les enseignants à adopter les modèles cognitifs qui les sous-tendent (Simon, Turcotte, & Forgette-Giroux, 2006). Les limites reliées à la mesure, comme la nature des échelles de questionnement et de notation, présentent des défis de taille lors des analyses secondaires des résultats à des fins pratiques et politiques. Ces défis ne se limitent pas nécessairement à la mesure mais comprennent également des difficultés techniques comme les techniques d'échantillonnage et pratiques comme l'accès aux données.

Lorsque les mesures s'avèrent valides, fidèles, et significatives, les résultats primaires et secondaires des programmes d'évaluation peuvent alors contribuer à l'établissement de plans de redressement fructueux. Les conclusions de deux recherches rapportant les résultats d'analyses secondaires laissent croire que, malgré les défis associés à ces analyses, certaines tendances s'observent et peuvent nourrir les politiques et les pratiques en matière d'apprentissage de la lecture et de l'écriture. Toutefois, il y a lieu d'encourager une plus grande collaboration entre chercheurs, praticiens, et dirigeants afin de rendre ces programmes d'évaluation davantage pertinents et utiles. Les études dans ce numéro spécial offrent des recommandations concrètes pouvant faire l'objet de telle collaboration.

such as the nature of item and scoring scales, pose major challenges for secondary analyses of results for practical and political purposes. Those challenges are not necessarily limited to the measurement instrument, but also include technical difficulties, such as sampling techniques, and practices, such as data access.

When tests are valid, reliable, and significant, the primary and secondary results of assessment programs can contribute to the development of meaningful corrective action plans. The findings of two studies involving the results of secondary analyses suggest that, despite the challenges associated with the analyses, certain trends can be observed and may be used in the development of policies and practices for learning to read and write. However, there is a need to encourage greater co-operation between researchers, practitioners, and administrators to make assessment programs more relevant and useful. This special issue offers specific recommendations for co-operation of that kind.

REFERENCES

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: AERA.

RÉFÉRENCES

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC : AERA.
- Bussière, P., Cartwright, F., Crocker, R., Ma, X., Oderkirk, J., & Zhang, Y. (2001). *À la hauteur : la performance des jeunes du Canada en lecture, en mathématiques et en sciences – Étude PISA de l'OCDE : premiers résultats pour les Canadiens de 15 ans*. Ottawa, ON : Développement des Ressources Humaines Canada, Statistique Canada.
- Bussière, P., Cartwright, F., Knighton, T., & Rogers, T. (2004). *À la hauteur : résultats Canadiens de l'étude PISA de l'OCDE : la performance des jeunes du Canada en mathématiques, en lecture, en sciences et en résolution de problèmes : premiers résultats de 2003 pour les Canadiens de 15 ans*. Ottawa, ON : Ressources Humaines et Développement des Compétences Canada, Statistique Canada, Conseil des Ministres de l'Éducation (Canada).
- Comité consultatif mixte. (1993). *Principes d'équité relatifs aux pratiques d'évaluation des apprentissages scolaires au Canada*. Edmonton, AB : Centre for Research in Applied Measurement and Evaluation.
- Joint Committee on Standards for Educational Evaluation. (2003). *The student evaluation standards*.
- Bussière, P., Cartwright, F., Crocker, R., Ma, X., Oderkirk, J., & Zhang, Y. (2001). *Measuring up: Performance of Canada's youth in reading, mathematics and science: OECD PISA study – First results for Canadians aged 15*. Ottawa, ON: Human Resources Development Canada, Statistics Canada.
- Bussière, P., Cartwright, F., Knighton, T., & Rogers, T. (2004). *Measuring up: Canadian results of the OECD PISA study – Performance of Canada's youth in mathematics, reading, science and problem-solving: First results in 2003 for Canadians aged 15*. Ottawa, ON: Human Resources and Skills Development Canada, Statistics Canada, Council of Ministers of Education, Canada.
- Joint Advisory Committee. (1993). *Principles for fair student assessment practices for education in Canada*. Edmonton, AB: Centre for Research in Applied Measurement and Evaluation.
- Joint Committee on Standards for Educational Evaluation. (2003). *Student Evaluation Standards*. Arlen Gullickson, Chair. Thousand Oaks, CA: Corwin.
- Landry, R., & Allard, R. (2002). *Résultats pancanadiens des élèves francophones en milieu minoritaire aux évaluations du PIRS : variables contextuelles et conséquences pédagogiques* (French only). Research report

- Arlen Gullickson, Chair. Thousand Oaks, CA : Corwin.
- Landry, R., & Allard, R. (2002). *Résultats pancanadiens des élèves francophones en milieu minoritaire aux évaluations du PIRS : variables contextuelles et conséquences pédagogiques*. Rapport de recherche soumis au Conseil des ministres de l'Éducation (Canada).
- Simon, M., Turcotte, C., & Forgette-Giroux, R. (2006). Impacts du PIRS en milieu scolaire. *Revue canadienne de l'évaluation de programme*, 21(1), 155–174.
- Wang, M.C., Haertel, G.D., & Walberg, H.J. (1993). Toward a knowledge base for school learning. *Review of Educational Research*, 63, 249–294.
- submitted to the Council of Ministers of Education, Canada.
- Simon, M., Turcotte, C., & Forgette-Giroux, R. (2006). Impacts du PIRS en milieu scolaire (French only). *Canadian Journal of Program Evaluation*, 21(1), 155–174.
- Wang, M.C., Haertel, G.D., & Walberg, H.J. (1993). Toward a knowledge base for school learning. *Review of Educational Research*, 63, 249–294.
- Marielle Simon**, a full professor, has been a member of the Faculty of Education at the University of Ottawa since 1994. Her research interests are in evaluation and assessment. She has obtained various SSHRC grants for secondary analyses of large-scale assessment data and classroom assessments.
- Marielle Simon** est professeure titulaire, membre de la Faculté de l'Éducation de l'Université d'Ottawa depuis 1994. Ses intérêts de recherche sont la mesure et l'évaluation. Elle a obtenu plusieurs subventions de CRSH dans le domaine de l'analyse secondaire des données des évaluations à grande échelle et de salle de classe.