

Does Ontario Have an Achievement Gap? The Challenge of Comparing the Performance of Students in French- and English-Language Schools on National and International Assessments

[Ruth Childs](#) & [Francine Dénoimé](#), OISE, University of Toronto

Abstract

On national and international assessments, students attending French-language schools in Ontario usually perform worse than students attending English-language schools. Interpreting these results is challenging because the French- and English-language schools differ both in prescribed curriculum and in how the curriculum is taught. In addition, the French- and English-language versions of the tests and scoring procedures sometimes differ. Even how students in the French- and English-language schools take the tests may differ. Finally, the populations of students differ in important ways. In this paper, we illustrate these challenges using results from the 2001 Progress in International Reading Literacy Study.

In the past decade, Ontario students have participated in numerous national and international assessments: the School Achievement Indicators Program (SAIP), assessing mathematics, reading, writing and science for 13- and 16-year-old students across Canada; the Trends in International Mathematics and Science Study (TIMSS), assessing mathematics and science for students in Grades 4 and 8; the Progress in International Reading Literacy Study (PIRLS), assessing reading for Grade 4 students; and the Programme for International Student Assessment (PISA), assessing reading literacy, mathematical literacy, and scientific literacy for 15-year-olds.

The results for Ontario students on the national and international assessments reveal a pattern: students attending French-language schools in Ontario usually perform worse than students attending English-language schools regardless of the subject area or grade level (Landry & Allard, 2002). In Ontario, French-language schools serve students with parents (1) whose first language is French, (2) who attended a French-language elementary school in Canada, (3) who have another child who is attending or has attended a French-language elementary or secondary school in Canada, and/or (4) who receive permission from an admissions committee (for example, because the grandparents' first language was French) (Ontario Ministry of Education and Training, 2004). French-language schools should not be confused with French immersion programs, which are administered by the English-language school boards and are intended for students who wish to learn French as a second language.

To understand the possible causes of this gap in achievement requires comparison of not only the assessment results, but also the prescribed curriculum for French- and English-language schools; what is actually taught, who teaches it, and how it is taught in the two school systems; the French- and English-language versions of the tests and the scoring procedures; students' test-taking behaviours; and the student populations. In this paper, we first review the research literatures in two areas of particular relevance for the comparison of language groups: the unique challenges that face minority language populations and the effects of test translation. We then illustrate these challenges using results from the 2001 PIRLS.

MINORITY LANGUAGE POPULATIONS

About 5% of Ontario elementary students attend French-language schools; at the secondary level, the percentage drops to 3%. Although they receive their instruction in French and may speak French at home, most of these students live in an English-speaking environment. Several recent studies have investigated the effects of minority language status on students' achievement.

Allen and Cartwright (2004), in a report entitled *Minority Language School Systems: A Profile of Students, Schools and Communities*, examined the lower achievement of French minority students in four Canadian provinces. The study used data from PISA 2000, on which French-language students did not perform as well as English-language students, and focused on the following three questions: (1) Are there other ways that students in French-language and English-language school systems differ? (2) Are there differences in the characteristics and resources of French-language and English-language schools? and (3) Are there other important differences in the families and communities of these students? Allen and Cartwright found that the French-language schools had fewer resources and that most Francophone students lived in predominantly English-speaking communities. The students attending the French-language schools were also less likely to speak at home the language in which the test was administered. Allen and Cartwright found no differences in socio-economic background between the two linguistic groups.

Not surprisingly, a survey of teachers in French-language schools in minority contexts across Canada reveals similar challenges. Gilbert, LeTouzé, Thériault, and Landry (2004) surveyed almost 700 teachers in such schools. The most often cited challenge was “living in French in an English-dominant setting,” followed by “lack of resources”; “lack of qualified staff” and “lack of physical facilities” were also identified as problems (Gilbert et al., 2004, pp. 27-29).

Gérin-Lajoie and Labrie (1999) investigated possible factors related to the performance of the Francophone students. Their study was commissioned by the association of teachers in Ontario French-language schools, l'Association des enseignantes et des enseignants franco-ontariens (AEFO), in response to the results of the SAIP 1993-1994 reading and writing assessment, on which Ontario students attending French-language schools performed worse than students attending English-language schools. As well as voicing serious concerns regarding the benefits of such assessments, Gérin-Lajoie and Labrie addressed the minority context, cultural differences within the Francophone community, linguistic skills and standards, and the possible impact of the marking procedures on the assessment results.

An additional challenge for French-language education in Ontario is the short time that French-language communities have governed their own schools. The current 12 French-language school boards (4 Public and 8 Separate) were established on January 1, 1998 and given the authority to manage their own schools. The publication by the Ontario Ministry of Education and Training in 2004 of the *Aménagement linguistique – A Policy for Ontario's French-Language Schools and Francophone Community* recognises the continuing challenges that face these young boards. The objectives of the policy include “deliver[ing] high-quality

instruction in French-language schools adapted to the minority setting” and “increasing the capacity of learning communities, including school staff, students, and parents, to support students' linguistic, educational, and cultural development throughout their lives” (Ontario Ministry of Education and Training, 2004, p. 4).

Other recent research has focused on minority populations more generally, whether or not defined by language. In October 2003, the Educational Testing Service published a document, *Parsing the Achievement Gap*, which identified correlates that create or perpetuate achievement gaps between minority and majority student populations. The three major categories of correlates were Early Development (e.g., birth weight, lead poisoning, and hunger and nutrition), School Environment (e.g., rigor of the school curriculum, teacher preparation, teacher experience and attendance, class size, and availability of appropriate classroom technology) and Home Learning Environment (e.g., reading to young children, watching television, parent availability and support, and student mobility). Individually, these correlates are not predictive; however, as clusters, they are the best researched predictors of achievement gaps between minority and majority student populations.

TEST TRANSLATION

While many researchers have focused on the characteristics of the students and the schools when trying to explain differences in achievement between populations, others have examined the test instruments themselves. In international assessments, test translation is a perpetual concern (Hambleton, 1993, 1994; Sireci, 1997). Assessments such as TIMSS, PISA, and PIRLS are written in English and/or French, and then translated into other languages. For example, PIRLS 2001 was prepared in English and then translated into 31 languages to be administered in 35 countries. In most international assessments, each participating country is responsible for translating the assessment, questionnaires, and other supporting materials into its language or languages of choice.

Most international assessments provide countries with guidelines for translating the assessments. For example, according to the *PIRLS 2001 Survey Operations Manual* (PIRLS, 2001b), there should be a minimum of two translators who at first work independently and then come together to arrive at one translated version. This version is then submitted to the central PIRLS committee, where it is checked by an independent translator who prepares a Translation Verification Report that includes recommended changes. The manual further advises that translators should pay particular attention to word equivalence, preserving word meaning, reading and difficulty level, ensuring correspondence between text in the passages and text in the items and lay-out modifications due to translation. In addition to these procedures and

guidelines, countries are provided with statistical analyses of item data from the field test and operational test administrations in order to check for evidence of differences in student performance that could be due to the translation. Countries are required to verify items that show unusual differences in item difficulty or patterns of distractor selection.

Even with all of these mechanisms in place to produce the most comparable assessment instrument possible, concerns regarding fairness, equity, and culture bias still exist. As Ellis (1989) reports, “even the most meticulous and painstaking translation and back translation will not ensure measurement equivalence” (p. 919). France’s Ministry of Education has been publicly critical of PISA, citing concerns regarding the impact of levels of difficulty in translated items, cultural biases, and the predominance of the Anglo-Saxon model used for the assessment (ÉduSCOL, 2003). Simon (1994) and Vaillancourt (1984) found that many items used in translated assessments were deemed to be biased when they were statistically analysed. Using studies such as the Second International Math Study (SIMS), Simon conducted differential item functioning analyses and found that approximately one third of the items functioned differently depending on the language of the test.

The *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999) includes an entire chapter on “Testing Individuals of Diverse Linguistic Backgrounds.” It states, “special attention to issues related to language and culture may be needed when developing, administering, scoring and interpreting test scores and making decisions based on test scores” (p. 91). The committee also warns that, “One cannot simply assume that such a translation produces a version of the test that is equivalent in content, difficulty level, reliability, and validity to the original untranslated version” (p. 92).

On assessments of reading literacy, in addition to the items themselves, the reading passages must be scrutinized. In most reading assessments, there are no indications that reading passages are subjected to any type of readability test to determine whether or not the translated passage is age- or grade-appropriate and of comparable difficulty.

The interpretation of the questionnaire items by the respondents must also be considered. Recent research by Simon, Turcotte, Ferne, and Forgette-Giroux (in press) suggests that teachers in Ontario’s French- and English-language schools differ in how they understand the PIRLS Teacher Questionnaire’s questions about classroom practices. These differences in interpretation may account for some of the differences in responses.

Finally, differences among communities in dialect and vocabulary are often ignored, but may cause the assessment materials to vary in difficulty across groups. This is particularly a

concern for minority language communities in Canada, where, for example, the French vocabulary may be different from that used in Quebec.

COMPARING ASSESSMENT RESULTS

The research on minority language populations and on test translation illustrates the range of possible influences on assessment results. The need to interpret test results in light of contextual factors that may influence students' opportunities to learn was recognized by researchers on the SIMS, who developed a three-part model (Travers & Westbury, 1989). According to this model, the National Social and Educational Context represents what society intends for students to learn and how the educational system should be organized. This is usually found in a variety of documents, such as the jurisdiction's official curriculum, guidelines, and policies. The School, Teacher, and Classroom Context represents what is actually taught in the classrooms, who teaches it, and how it is taught. The third aspect is the Student Outcomes and Characteristics, which corresponds with what the students have actually learned and their attitudes regarding the subjects.

While this model acknowledges the importance of understanding differences in curricula and school practices, it does not include characteristics of the assessment instruments, such as possible translation effects. In this study, the characteristics of the assessment instruments, how students interact with those instruments, and the scoring procedures will all be considered, in addition to the three aspects in the model. We will refer to this aspect simply as the Assessment and place it between the original second and third aspects.

METHOD

Participants

Thirty-five countries participated in PIRLS 2001. In Canada, only the provinces of Ontario and Quebec participated. PIRLS uses a two-stage stratified cluster sample design where schools are selected, then one classroom from the grade with the majority of 10-year-old students (in Ontario, Grade 4) is selected within each school. In Ontario, private, Aboriginal, special needs, and very small schools (fewer than 10 students in Grade 4) were excluded from the sample. Ontario's sample design included explicit stratification by language (French and English) and school size (large and very large schools).

In total, 122 English schools and 80 French schools were sampled in Ontario in order to collect sufficient data for both language groups. From the sample, 116 of the 122 selected English-language schools (95%) and 74 of the 80 selected French-language schools (93%) participated. From the Grade 4 classroom selected within each school, all students were

expected to participate unless they belonged to one of the following groups: educable mentally disabled students, functionally disabled students or non-native-language speakers. Over 4,000 Ontario students participated in PIRLS 2001: approximately 1,500 were Francophone students and 2,700 were Anglophone students. In the analyses of students' responses, each student received a weight proportional to the inverse of that student's probability of being selected; these weights were used to correct for the different selection probabilities of schools and classrooms (Joncas, 2003).

Instruments

The PIRLS, established in 1998 by the International Association for the Evaluation of Educational Achievement (IEA), was administered for the first time in 2001 and will be administered at five-year intervals. Its aim is to investigate children's literacy skills and factors associated with the acquisition of those skills. The PIRLS contains two types of reading passages: literary and informational. In PIRLS 2001, each student received an 80-minute booklet containing several reading passages. There were 10 booklets and some passages appeared in more than one booklet. Campbell, Kelly, Mullis, Martin, and Sainsbury (2001) provide examples of reading passages for each purpose and their accompanying items. For example, one of the Reading for Literary Purposes passages, "The Dressmaker," is a short story about a retiring tailor who passes on his sewing machine and business to a young girl. It is accompanied by eight multiple-choice items and four constructed-response items (each scored on a two- or three-point scale). One of the Reading for Informational Purposes passages, "Puppy Walking," describes how a family helps train a puppy to become a guide dog; it is accompanied by seven multiple-choice items and six constructed-response items. For both types of passages, the accompanying items are intended to measure four comprehension processes: (1) Focus on and Retrieve Explicitly Stated Information, (2) Make Straightforward Inferences, (3) Interpret and Integrate Ideas and Information, and (4) Examine and Evaluate Content, Language, and Textual Elements.

Ontario and Quebec collaborated in scoring their students' assessments. The English version of the assessment was scored in Ontario by 30 scorers from Ontario and 10 scorers from Quebec. The French version was scored in Quebec by 30 Quebec scorers and 10 Ontario scorers. All markers were either current or retired teachers. Each marker received a copy of the *PIRLS Scoring Guides for Constructed-Responses* (PIRLS, 2001a), which they were instructed to follow precisely to score the constructed-response items. The guides included anchor papers (examples of student responses at particular score levels), and practice papers (pre-scored papers intended to help markers achieve accuracy and consistency in scoring).

The PIRLS 2001 used four questionnaires to collect information on factors expected to be associated with students' literacy achievement (Kelly, 2003). Each student completed a 30-minute Student Questionnaire about their attitudes toward reading, their reading activities, and the literacy resources in their home. Principals completed a 30-minute School Questionnaire about the characteristics of the school and students, the literacy curriculum, and the school's literacy resources. The homeroom teacher of the sampled classroom answered a 30-minute Teacher Questionnaire about the size and other characteristics of the class, the literacy resources in the classroom, and his or her instructional and assessment activities and professional training. Finally, parents or caregivers were asked to complete a 15-minute Learning to Read Survey providing information about their child's early language and literacy experiences, the parents' reading attitudes and activities, and the literacy resources in their home. All of the questionnaire items required respondents to select from provided responses.

Analyses

The National Social and Educational Context

Responses of the school principals to questions on the School Questionnaire about the level of preparation with which students entered their schools were compared. Chi-square tests were used to determine whether the distributions of responses across the four response categories ("less than 25%," "25-50%," "51-75%," and "more than 75%") were significantly different between the English- and French-language principals.

The School, Teacher, and Classroom Context

Teachers' and principals' responses to questions related to the school, teacher, and classroom context on the Teacher Questionnaire and School Questionnaire, respectively, were also compared. It is important to note that because the schools were sampled and not the teachers, the teachers who responded to the Teacher Questionnaire are not a representative sample of Grade 4 teachers within Ontario; they were simply the teachers who taught the students in the classrooms that were selected to participate in the study. The teachers' responses were therefore weighted inversely to the sampling probability of the school and classroom, so that the teachers' responses can be assumed to represent the responses of teachers for a representative sample of students in Ontario. The *PIRLS 2001 User Guide for the International Database* (Campbell et al., 2003) warns that it is only appropriate to make statements about the teachers in terms of how many students are taught by teachers who provided particular responses. Similar caveats, of course, apply to the principals' responses.

The Assessment

The translation of items from English to French was examined to investigate the possibility that item difficulty had been affected or that the meaning of the item had been influenced. The French-language and English-language scoring guides were also compared and the scoring schemes were examined. How the marking session was conducted was considered for possible cultural and marker bias. These comparisons were performed by the first author, who is fluently bilingual and was Ontario's principal liaison for the PIRLS 2001.

Analysis of the students' responses to the assessment items provided an indication of how the students interacted with the assessment. The PIRLS 2001 Almanacs, provided to all participating countries, were used in this analysis. The Almanacs provided correlations for the items found in all four background questionnaires with average student achievement scores. They also provided classical item analysis results for all of the items of the assessment for both the French- and English-language students in Ontario. The data included the number of participants, the difficulty index, the discrimination index, the percentage-correct, the percentage of students who did not reach the item and the percentage of students who omitted the item.

Student Outcomes and Characteristics

Finally, students' scores on the PIRLS 2001 were compared, as evidence of the Attained Curriculum. Students' responses to items in the Student Questionnaire, particularly related to their attitudes toward reading, were also analyzed.

RESULTS AND DISCUSSION

The National Social and Educational Context

Table 1 presents French- and English-Language principals' responses regarding students' prior knowledge and experience as they enter Grade 1. *The Report of the Expert Panel on Early Reading in Ontario* (Ontario Ministry of Education, 2003) defines prior knowledge and experience as "the world of understanding that children bring to school" (p. 15). In all cases, principals reported that English-language students enter Grade 1 in the English-language schools with more skills and knowledge than students in the French-language schools. Principals of 62% of English-language students reported that more than 75% of their students begin Grade 1 with the ability to recognize most of the letters of the alphabet. The percentage is much lower in the French schools: principals of less than 35% of French-language students reported that more than 75% of their students have this skill as they begin Grade 1. For all of

the skills and knowledge presented in Table 1, principals for many more French-language students indicated that less than 25% of their students have acquired some of the basic skills.

The distributions of responses of the English- and French-language principals were significantly different regarding the percentages of their students who entered Grade 1 being able to recognize most letters of the alphabet, $\chi^2(3) = 42.879, p < .001$; read some words, $\chi^2(3) = 40.690, p < .001$; read some sentences, $\chi^2(3) = 25.013, p < .001$; write letters of the alphabet, $\chi^2(3) = 36.438, p < .001$; and write some words, $\chi^2(3) = 45.700, p < .001$.

The School, Teacher, and Classroom Context

In the PIRLS 2001 Teacher Questionnaire, teachers were asked to describe their instructional practices, use of resources, and assessment practices. For example, Table 2 presents the frequency of use of different types of assessment tools by teachers to monitor student performance. Teachers reported that more Francophone students are exposed at least once a month to multiple-choice questions (68.3%) than Anglophone students (51.7%). More Anglophone students, however, are asked at least once a week to use short-answer responses (55.4%) and paragraph-length responses (39.0%) than Francophone students (35.1% and 15.6%, respectively). The use of oral questioning of students, asking students to give an oral summary or report of what they have read and meeting with students to discuss what they have been reading and work they have done are assessment strategies and tools that are more frequently used by English-language teachers than by French-language teachers. A higher percentage of Francophone students than Anglophone students are reported as never being exposed to these strategies.

There were significant differences between the French-language and English-language teachers' responses in the frequency of the use of some assessment strategies and tools to monitor students' progress in reading: short-answer written questions on material read, $\chi^2(2) = 10.188, p < .01$; paragraph-length written responses about what students have read, $\chi^2(3) = 22.747, p < .001$; listening to students read aloud, determining oral reading accuracy, $\chi^2(3) = 9.970, p < .05$; oral questioning of students, $\chi^2(3) = 16.769, p < .01$; students give an oral summary/report of what they have read, $\chi^2(3) = 9.708, p < .05$; meeting with students to discuss what they have been reading and work they have done, $\chi^2(3) = 31.465, p < .001$. The differences were not significant for other strategies and tools: multiple-choice questions on material read, $\chi^2(3) = 6.310, p = .097$; $\chi^2(3) = 6.909, p = .075$; listening to students read aloud, $\chi^2(3) = 6.909, p = .075$.

The Assessment

Differences in language difficulty on the French and English versions of the assessment were investigated. For example, the item in Table 3 accompanies the literary text “The Upside-Down Mice” by Roald Dahl, 1981. A word-for-word translation of the English version into French would be “Quels mots décrivent mieux cette histoire?” A translation of the French version into English would be “Which adjectives best describe this story?” The more precise, but less familiar word “adjectives” instead of “words” likely increases the difficulty level of this item. In the options, the words “scary,” “clever,” and “thrilling” were translated as “effrayante,” “ingénieuse,” and “palpitante.” These words are not as common as the adjectives found in the English version and also likely increase the difficulty level of this item. The options acted differently in the English and French version of the test, as more than twice as many Francophone students than Anglophone students chose option A. Options B and D were chosen approximately three times more often by Francophone students than by Anglophone students.

The concerns regarding the translation of the English texts and items to French also apply to the scoring guides developed for the constructed-response items. The scoring guides include the following elements: the purpose and process, the question, the score point attributed, and response description, a detailed explanation of the response, evidence, and examples. The purpose relates to the reason why people read: for literary experience or to acquire and use information. PIRLS 2001 assesses four types of comprehension processes: Focus on and Retrieve Explicitly Stated Information; Make Straightforward Inferences; Interpret and Integrate Ideas and Information; and Examine and Evaluate Content, Language, and Textual Elements. The scoring guide indicates the number of points and a response description, such as Complete Comprehension, Partial Comprehension, No Comprehension, Acceptable Response, or Unacceptable. A detailed explanation of the response follows, providing the essential elements of the answer. Finally, examples are provided. The examples are authentic student responses taken from English responses. All elements of the scoring guides are translated from English to French. These include the student responses, which means that there are no authentic responses from the French cohort. The responses are grammatically correct and contain no spelling errors.

The one-point item in Table 4 also refers to the literary text “The Upside-Down Mice.” Results for this question are as follows: 50.8% of Anglophone students and 33.9% of Francophone students received 1 point, 44.9% of Anglophone students and 57.5% of Francophone students received a score of 0. This question was omitted by 4.3% of Anglophone students and by almost twice as many Francophone students, 8.5%.

The first obvious difference between the French and English versions is the description of the purpose in the scoring guide. In the English version, the purpose is said to be “Literary,” while in the French version it is described as a “Reading Experience,” “une expérience de lecture.” Other differences in the scoring guide are found in the description of what is an acceptable response or “une réponse acceptable.” Whereas the word “appropriate” is used to describe the type of interpretation provided by the Anglophone student to receive a score of 1 point, the French version uses the word “juste” or a “just interpretation.” It is questionable whether the English-language scorers and the French-language scorers would interpret the words “appropriate” and “juste” in the same way. The English version also includes the word “whole” in the description: “These responses provide an appropriate interpretation of Labon’s reaction within the context of the whole story.” In the French version, the word “whole” is omitted: “dans le contexte de l’histoire.”

The English and French versions of the PIRLS 2001 were scored in separate sessions by different markers. The consistency of the training across these scoring sessions is not known.

Information about students’ responses to individual items was also analyzed. For example, Table 5 presents the student results on the three-point constructed-response items, which are the most complex of the test items. The Anglophone students performed better than the Francophone students for all seven of the three-point items. More Francophone students omitted or failed to reach each of these items.

Student Outcomes and Characteristics

Results of the PIRLS 2001 are reported for Overall Reading Achievement, Achievement in Reading for Informational Purposes and Achievement in Reading for Literary Purposes. Students’ performance is expressed as a score on a scale from 0 to 1000, with an international average of 500.

Ontario Grade 4 students achieved an average score of 548 in Overall Reading Achievement, 551 in Reading for Literary Purposes, and 542 in Reading for Informational Purposes. Three countries performed significantly better: Sweden (559), the Netherlands (553), and Bulgaria (551).

As was stated earlier, when Ontario’s Grade 4 population is broken down by language, Ontario’s Anglophone students’ average performance does not change with a score of 550 but the Grade 4 Francophone students scored significantly lower at 494 for Overall Reading Achievement, 488 for Reading for Literary Purposes and 501 for Reading for Informational

Purposes. Without this breakdown, the performance of the minority Francophone students in the province of Ontario is masked by the performance of the majority Anglophone students. The Ontario Francophone students performed significantly below the international average, with only 8 countries of the 35 participating countries performing significantly lower. Quebec's Francophone students, who completed the same version as Ontario's Francophone students, performed significantly better than Ontario Francophone students, with a score of 537, but significantly worse than Ontario's Anglophone students. Quebec's Anglophone students, who wrote the same version as Ontario's Anglophone students performed similarly to Ontario's Anglophone students, with a score of 543.

CONCLUSION

These analyses illustrate some of the challenges of interpreting differences in performance on large-scale assessments. As the results show, there are important differences in the educational experiences of the students and in the versions of the assessments they write. For example, fewer students entering French-language schools have preliteracy skills when they begin their formal schooling and fewer of these students receive practice providing written responses to what they read.

When we compare the performance of students in the French- and English-language schools on individual items, it is clear that the students responded very differently to many of the items. Some of these differences may be due to differences in preliteracy skills or in the teachers' instructional and assessment practices. However, comparison of the item texts suggests that some of these differences may be due to differences in meaning or difficulty introduced during translation. Examination of the scoring guides also revealed differences in meaning due to translation.

There is a need for many more studies to provide French-language educators and policy-makers with the information they need to improve French-language education in Ontario. For example, Simon et al.'s (in press) recent study suggests that, before we can make recommendations about teachers' classroom practices based on the results of the PIRLS Teacher Questionnaire, more research is needed to understand how teachers are interpreting and responding to the questions on such questionnaires. The effect of choice of vocabulary in translating the materials for students living in minority versus majority French communities should be studied. The results for students who begin school speaking French versus those who do not should be compared.

In conducting these and other studies using data from large-scale assessments, such as PIRLS, we urge researchers to create a comprehensive picture of the experiences of students

and the characteristics of the assessments. Assuming that translation differences do not exist or that resources and instructional practices are similar across schools can easily lead to misleading conclusions. We owe better to the students and teachers.

REFERENCES

- Allen, M., & Cartwright, F. (2004). Minority language school systems: A profile of students, schools and communities. *Education Quarterly Review, 9*, 9-48.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Campbell, J. R., Foy, P., Gonzalez, E.J., Hastedt, D., Itzlinger, U., Joncas, M., Kelly, D. L., Malak, B., Martin, M. O., Mullis, I. V. S., Sainsbury, M. & Schwippert, K. (2003). *PIRLS 2001 User guide for the international database*. Boston: International Study Center, Lynch School of Education, Boston College.
- Campbell, J. R., Kelly, D. L., Mullis, I. V. S., Martin, M. O., & Sainsbury, M. (2001). *Framework and specifications for PIRLS assessment 2001*. Boston: International Study Center, Lynch School of Education, Boston College.
- ÉduSCOL. (2003). *Évaluations des connaissances et des compétences des élèves de 15 ans. Le cas français. Ministère de la Jeunesse, de l'Éducation nationale et de la Recherche*. Paris: Direction de l'Enseignement scolaire.
- Ellis, B. B. (1989). Differential item functioning: Implications for test translations. *Journal of Applied Psychology, 74*, 912-920.
- Gérin-Lajoie, D., & Labrie, N. (1999). Les résultats aux tests de lecture et d'écriture en 1993-1994 : une interprétation sociolinguistique. Dans N. Labrie et G. Forlot (Eds.), *L'enjeu de la langue en Ontario français* (pp. 79-108). Sudbury, ON: Prise de parole.
- Hambleton, R. K. (1993). Translating achievement tests for use in cross-cultural studies. *European Journal of Psychological Assessment, 9*, 57-68.
- Gilbert, A. LeTouzé, S. Thériault, J. Y., & Landry, R. (2004). *Teachers and the challenge of teaching in minority settings: Final research report*. Ottawa: Canadian Teachers' Federation.
- Hambleton, R. K. (1994). Guidelines for adapting educational and psychological tests: A progress report. *European Journal of Psychological Assessment, 10*, 229-244.
- Joncas, M. (2003). PIRLS sampling weights and participation rates. In M. O. Martin, I. V. S. Mullis, & A. M. Kennedy (Eds.), *PIRLS 2001 technical report* (pp. 113-133). Boston: International Study Center, Lynch School of Education, Boston College.
- Kelly, D. L. (2003). Developing the PIRLS Background Questionnaires. In M. O. Martin, I. V. S. Mullis, & A. M. Kennedy (Eds.), *PIRLS 2001 technical report* (pp. 29-37). Boston: International Study Center, Lynch School of Education, Boston College.

- Landry, R., & Allard, R. (2002, octobre). *Résultats pancanadiens des élèves francophones pédagogiques*. Toronto: Le Conseil des ministres de l'Éducation.
- Ontario Ministry of Education and Training. (2003). *Early reading strategy – The report of the expert panel on early reading in Ontario*. Toronto: Queen's Printer for Ontario.
- Ontario Ministry of Education and Training. (2004). *Aménagement linguistique – A policy for Ontario's French-language schools and Francophone community*. Toronto: Queen's Printer for Ontario.
- Progress in International Reading Literacy Study (PIRLS). (2001a). *Scoring Guides for the Constructed-Response Items* (PIRLS Ref. No. 01-0007). Chestnut Hill, MA: Boston College.
- Progress in International Reading Literacy Study (PIRLS). (2001b). *Survey Operations Manual* (PIRLS Ref. No. 01-0001). Chestnut Hill, MA: Boston College.
- Simon, G. M. (1994). Differential item functioning: Applicability in a bilingual context. In Lavault, D. Zumbo, B. D., Bessaroli, M. E. & Boss, M. W. (Eds.) *Modern theories of measurement: Problems and issues*. Ottawa, ON: Faculty of Education, University of Ottawa.
- Simon, M., Turcotte, C., Ferne, T., & Forgette-Giroux, R. (sous presse). Pratiques pédagogiques dans les écoles de langue françaises de l'Ontario selon les données contextuelles du PIRLS 2001. *Mesure et évaluation en éducation*.
- Sireci, S. G. (1997). Problems and issues in linking assessments across languages. *Educational Measurement: Issues and Practices*, 16, 12-19.
- Travers. K. J., & Westbury, I. (1989). *The IEA Study of Mathematics I: Analysis of Mathematics Curricula*. Oxford: Pergamon Press.
- Vaillancourt, R. (1984). *IRT bias detection techniques compared with classical item analysis as applied to Anglophones and Francophones*. Paper presented at the annual meeting of the National Council for Measurement in Education, New Orleans.

Table 1

Responses of Principals in French- and English-Language Schools Regarding Students' Readiness to Learn

Percentage of Students						
Who Can Do the Following						
When They Begin their First Year of Formal Schooling (Grade 1)	Language of School	N	Less than			More than
			25%	25-50%	51-75%	75%
Recognize most of the letters of the alphabet	English	111	6.0	7.2	24.9	62.0
	French	72	43.4	12.6	9.3	34.7
Read some words	English	111	10.1	20.8	30.8	38.4
	French	72	51.8	9.5	11.5	27.2
Read some sentences	English	111	36.9	32.6	25.4	5.0
	French	72	72.4	7.1	16.7	3.9
Write letters of the alphabet	English	111	7.8	10.3	28.8	53.2
	French	71	43.0	15.1	12.6	29.3
Write some words	English	111	17.3	21.9	26.9	33.8
	French	72	66.5	8.3	13.0	12.2

Table 2***Responses of Teachers in French- and English-Language Schools Regarding the Use of Assessment Strategies and Tools to Assess Students' Performance in Reading***

Assessment Strategies and Tools	Language of School	<i>N</i>	At least once a week	Once or twice a month	Once or twice a year	Never
Multiple-choice questions on material read	English	128	7.1	44.6	31.7	16.7
	French	80	10.0	58.3	24.0	7.8
Short-answer written responses on material read	English	129	55.4	40.3	4.3	0.0
	French	80	35.1	62.9	2.0	0.0
Paragraph-length written responses about what students have read	English	128	39.0	48.7	8.2	4.2
	French	80	15.6	48.2	28.6	7.5
Listening to students read aloud	English	129	56.1	36.4	7.5	0.0
	French	80	64.0	27.5	4.8	3.6
Determining oral reading accuracy	English	125	38.3	39.2	20.4	2.1
	French	79	28.4	41.0	18.2	12.4
Oral questioning of students	English	128	75.4	21.6	2.1	0.9
	French	80	50.2	35.9	8.7	5.3
Students give an oral summary/report of what they have read	English	129	33.3	43.4	20.3	2.9
	French	80	17.1	44.0	30.9	8.0
Meeting with students to discuss what they have been reading and work they have done	English	128	23.8	54.1	18.7	3.5
	French	79	11.4	28.5	44.2	15.9

Table 3

A Multiple-Choice Literary Item with More than 30% Difference in Percentage of Students in French- and English-language Schools Answering Correctly

Item Stem	Options			
	A	B	C*	D
Which words best describe this story?	Serious and sad 6.4%	Scary and exciting 2.7%	Funny and clever 82.0%	Thrilling and mysterious. 7.4%
Quels adjectifs décrivent le mieux cette histoire?	Elle est sérieuse et triste. 14.3%	Elle est effrayante et excitante. 10.6%	Elle est amusante et ingénieuse. 50.5%	Elle est palpitante et mystérieuse. 21.6%

Table 4

French and English Scoring Guides for a One-Point Constructed-Response Item

English Version	French Version
Purpose: Literary	But : Expérience de lecture
Process: Interpret and Integrate Ideas and Information	Processus : Interpréter et assimiler des idées et de l'information
Question: Why did Labon smile when he saw there were no mice in the traps?	Question : Pourquoi M. Labon sourit-il en voyant les pièges vides?

English Version	French Version
<p>1 point – Acceptable Response</p> <p>These responses provide an appropriate¹ interpretation of Labon’s reaction within the context of the whole² story.</p> <p>Evidence:</p> <p>The response demonstrates understanding that Labon was not surprised by the empty traps. It may describe Labon’s intent to carry out a more elaborate plan for catching the mice.</p> <p>Examples:</p> <p><i>-1 He had a plan to fool the mice and get rid of them.</i></p> <p><i>-2 Because he had other things in mind for the mice.</i></p> <p>Or, it may demonstrate understanding that he had intended to fool the mice, not to catch them, on the first night.</p> <p>Examples:</p> <p><i>-1 He knew that they would not go for the cheese the first night.</i></p> <p><i>-2 He had fooled the mice into thinking he was stupid.</i></p>	<p>1 point – Réponse acceptable</p> <p>Ces réponses donnent une juste¹ interprétation de la réaction de M. Labon dans le contexte de l’histoire.</p> <p>Preuve:</p> <p>La réponse montre que l’élève a compris que M. Labon n’est pas surpris de trouver les pièges vides. Elle peut indiquer que M. Labon a l’intention de mettre à exécution un plan plus élaboré pour attraper les souris.</p> <p>Exemples:</p> <p><i>-1 Il veut tromper les souris et s’en débarrasser.</i></p> <p><i>-2 Parce qu’il a d’autres choses en tête pour les souris.</i></p> <p>La réponse peut aussi montrer que l’élève a compris que, la première nuit, M. Labon a seulement l’intention de tromper les souris et non pas de les attraper.</p> <p>Exemples:</p> <p><i>-1 Il sait qu’elles n’iront pas chercher le fromage la première nuit.</i></p> <p><i>-2 Il s’arrange pour que les souris le croient stupide.</i></p>

English Version	French Version
<p>0 – Unacceptable Response</p> <p>These responses do not provide an appropriate¹ interpretation of Labon’s reaction within the context of the whole² story.</p> <p>Evidence:</p> <p>The response includes no evidence of understanding that the empty traps were what Labon expected to find, or that he intended to carry out a more elaborate plan for catching the mice. The response may simply restate his reaction without providing an appropriate¹ interpretation for it.</p> <p>Non-Response Codes</p> <p>8 – Not administered. Question misprinted, page missing, or other reason out of student’s control.</p> <p>9 – Blank</p>	<p>0 – Réponse inacceptable</p> <p>Ces réponses ne donnent pas une juste¹ interprétation de la réaction de M. Labon dans le contexte de l’histoire.</p> <p>Preuve:</p> <p>La réponse ne montre pas que l’élève a compris que M. Labon s’attend effectivement à trouver les pièges vides ou qu’il a l’intention de mettre en œuvre un plan plus élaboré pour attraper les souris. Elle peut simplement exposer de nouveau la réaction de M. Labon sans en donner une interprétation juste¹.</p> <p>Aucune réponse – Codes</p> <p>8 – Partie du test non administrée. Question mal imprimée, page manquante ou toute autre raison indépendante de la volonté de l’élève.</p> <p>9 – Blanc</p>

Table 5
Scores of Students in French- and English-language Schools on Three-Point
Constructed-Response Items

ITEM	Language of School	N					Not	
			3 points	2 points	1 point	0 point	reached	Omitted
R011M12 C	English	675	18.7	30.7	28.2	13.3	0.7	8.4
	French	398	11.1	24.9	27.4	24.1	1.5	11.1
R011C10 C	English	672	39.7	18.3	23.2	11.6	0.6	6.6
	French	381	26.5	16.5	20.7	23.9	1.8	10.5
R011A07 C	English	688	55.2	24.3	11.6	7.3	0.4	1.2
	French	384	47.1	14.8	18.2	15.1	0.0	4.7
R011L04 C	English	684	15.4	29.1	38.5	12.9	0.0	4.2
	French	376	9.3	9.8	59.0	13.0	0.0	8.8
R011R10 C	English	687	11.4	43.7	25.0	11.9	1.3	5.8
	French	385	3.1	47.5	26.0	14.0	2.6	6.8
R011R11 C	English	687	18.8	37.7	25.6	11.9	3.1	2.9
	French	385	16.6	29.6	32.2	9.1	6.2	6.2
R011H10 C	English	709	13.4	45.4	14.1	21.3	1.0	4.8
	French	376	6.7	41.2	11.4	26.3	4.0	10.4

