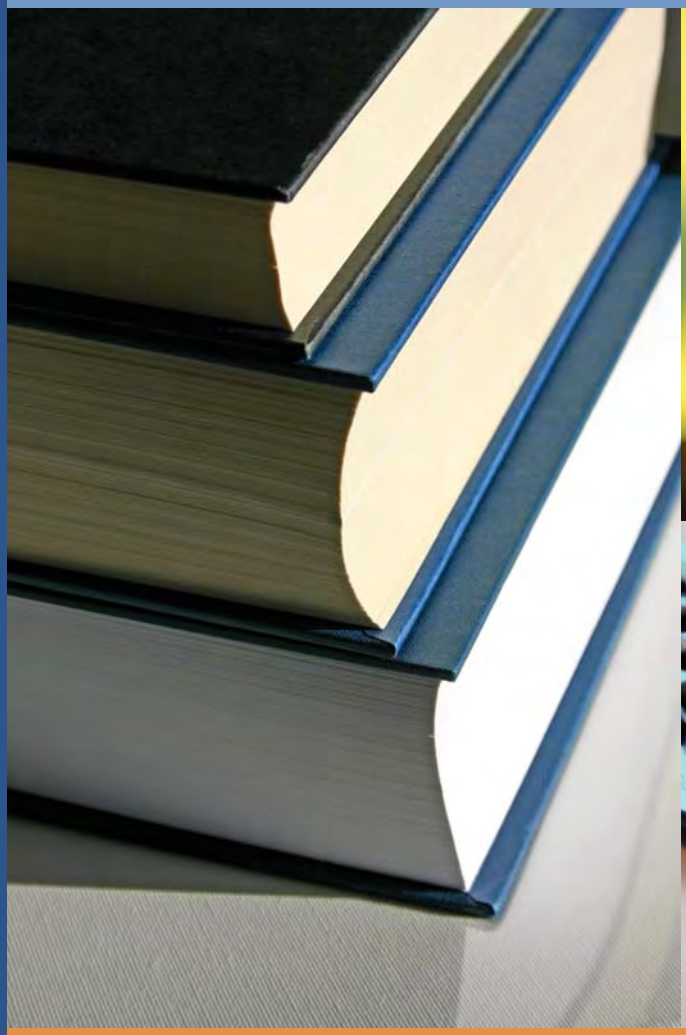
Pan-Canadian Assessment Program

PCAP-13 2007

Jurisdictional Profiles and Achievement Equity



cmec
Council of Ministers of Education, Canada
Conseil des ministres de l'Éducation (Canada)

Pan-Canadian Assessment Program

PCAP-13 2007

Jurisdictional Profiles and Achievement Equity

Developed by

Robert Crocker
Victor Glickman

cmec
Council of Ministers of Education, Canada
Conseil des ministres de l'Éducation (Canada)

The Council of Ministers of Education, Canada (CMEC) was formed in 1967 by the jurisdictional ministers responsible for education to provide a forum in which they could discuss matters of mutual interest, undertake educational initiatives cooperatively, and represent the interests of the provinces and territories with national educational organizations, the federal government, foreign governments, and international organizations. CMEC is the national voice for education in Canada and, through CMEC, the provinces and territories work collectively on common objectives in a broad range of activities at the elementary, secondary, and postsecondary levels.

Through the CMEC Secretariat, the Council serves as the organization in which ministries and departments of education undertake cooperatively the activities, projects, and initiatives of particular interest to all jurisdictions[1]. One of the activities on which they cooperate is the development and implementation of pan-Canadian testing based on contemporary research and best practices in the assessment of student achievement in core subjects.

Note of appreciation

The Council of Ministers of Education (Canada) would like to thank the students, teachers, and administrators whose participation in the Pan-Canadian Assessment Program ensured its success. The quality of your commitment has made this study possible. We are truly grateful for your contribution to a pan-Canadian understanding of educational policy and practices in reading, mathematics, and science among 13-year-olds.

Council of Ministers of Education, Canada
95 St. Clair West, Suite 1106
Toronto, Ontario M4V 1N6

Telephone: (416) 962-8100
Fax: (416) 962-2800
E-mail: cmec@cmec.ca

ISBN 978-0-88987-226-4

Ce rapport est également disponible en français.

Printed on recycled paper.

[1] In this report, “ministry” includes “department” and “jurisdictions” includes participating “provinces” and “territories.”

EXECUTIVE SUMMARY

The Pan-Canadian Assessment Program (PCAP) is CMEC's most recent commitment to informing Canadians about how well their education systems are meeting the needs of students and society. The first PCAP administration took place in 2007 and assessed the performance of 13-year-old students in reading, mathematics, and science. The PCAP samples were selected to allow reporting by jurisdiction (province and territory) and by official-language grouping within jurisdictions. Detailed reports on PCAP 2007 are available from CMEC (CMEC, 2008; 2009).

This report is part of a projected series of research projects in which the PCAP database is used to examine questions of interest to educational policy-makers in Canada. This report focuses on differences between jurisdictions and language groups (collectively referred to as populations in PCAP) and on factors contributing to these differences. The existence of differences raises the further questions of equity in educational experiences across and within populations.

The conceptual framework for this report is an educational productivity model, set within a human capital theory framework. This model takes school achievement as the main outcome of interest. Increasing average achievement and reducing disparities in achievement are treated as important goals for educational policy-makers. Studies such as this are intended to contribute to the evidence needed for policy-making and policy change in Canada.

The report begins with a brief review of literature on factors contributing to school achievement, focusing explicitly on studies of jurisdictional differences and based mainly on results from large-scale international assessments. Generally, the results show that differences within Canadian jurisdictions are larger than those between. Many factors contribute to jurisdictional differences, but none seems to be decisive. Important factors include socioeconomic status at the jurisdiction level, some large-scale reform policies, and teaching quality. Resource levels, in themselves, seem to have little effect. Private schools and school autonomy have positive effects on achievement, but negative effects for equity are found for highly differentiated systems. More generally, high average achievement and a high degree of equity are not conflicting goals. Some countries, including Canada, have managed to combine high achievement with relatively low differences between the highest and lowest-performing students.

The second stage of the report examines historical patterns of jurisdictional and language differences in Canada, using results derived from close to two decades of national and international assessments yielding results at the jurisdictional and official-language levels. The results show that differences across jurisdictions and languages are of the order of 0.60 to 0.70 standard deviation units or 20%–30% of students meeting acceptable standards (depending on how the results were reported). These differences have been highly persistent over time, with students in Alberta and Quebec consistently showing the highest performance and those in the Atlantic provinces the lowest. Francophone students

in minority settings (all jurisdictions outside Quebec) tend to perform at lower levels than their anglophone counterparts, whereas the two language groups within Quebec show little difference.

The analytical stage of the study was based on reading scores from the PCAP-13 2007 assessment. Analytical challenges in attempting to model population (jurisdiction and language) differences led to a two-stage approach. In the first stage, three sets of "jurisdictional profiles" were developed, based on student, teacher, and school variables shown to be significantly associated with reading achievement and showing significant differences across jurisdictions. The results show that the PCAP populations[2] are more alike than different on most of the variables included in the profiles. This suggests that the available variables are not likely to have significant explanatory value. Differences between the two jurisdictions with the highest and lowest reading scores for each language group (Ontario and Prince Edward Island for English and Quebec and New Brunswick for French) did reveal some patterns of interest. These are listed here. However, the details in the report should be read to avoid the risk of over-interpreting these results.

- Both of the highest-performing populations (Ontario English and Quebec French) have more students in Grade 8 (Secondary 2 in Quebec) than the lowest-performing populations (Prince Edward Island English and New Brunswick French). Both of the latter actually have more students in Grade 9 than any other population.
- Both of the highest-performing populations have more students born outside of Canada than the lowest-performing populations.
- Teachers in the highest-performing populations expect more homework from their students than in low-performing populations, and students in the highest-performing populations do more homework than those in the lowest-performing populations.
- Quebec francophones have more books in their homes than New Brunswick francophones.
- More students in Ontario English than in Prince Edward Island English know what a scoring rubric is.
- Both of the high-performing populations have higher percentages of classes with more than 30 students than the low-performing populations.
- Teachers in Prince Edward Island English use homework for grading more often than those in Ontario English.
- Quebec francophone teachers more often re-teach basic reading skills than their New Brunswick francophone counterparts.
- Higher teacher assessment skill is positively related to achievement. Ontario English teachers reported having higher levels of assessment skill than Prince Edward Island English.
- Using non-academic criteria for grading occurs much more often for Prince Edward Island English teachers than for either their Ontario English counterparts or those in either of the francophone populations.

[2] The term "population" is used throughout this report to refer to the combination of jurisdictions (provinces and territories) and official-language groups (English and French). This term reflects the fact that samples for pan-Canadian assessments (SAIP and PCAP) are drawn from these population groupings.

- More schools in Ontario English and Quebec French, the high-performing populations, are larger and are in larger communities than in the low-performing populations.
- Students in private schools, especially in Quebec, have higher reading achievement than those in public schools, even after controlling for other variables. The private school effect is a plausible factor in accounting for the exceptionally high performance of Quebec francophone students.
- New Brunswick francophone schools are more likely to spend more than 300 minutes per week on language arts than is the case for other populations.

The second stage of analysis involved the use of hierarchical modelling to examine the relative contribution to population differences in achievement of each of the variables used in the previous section, controlling for all of the other variables. Although the structure of the data lends itself to a three-level model (students within schools within populations), there are major analytical difficulties in estimating such a model with a small number of populations. The approach taken was thus to build an initial "population" model, which gives a coefficient for each population, representing the differences in average reading score between the population and a "reference population" (Ontario English or Quebec French). A series of intermediate models and finally a "full model" were then computed, including the populations and all of the other variables. Changes in the coefficient for each population, relative to its value in the population model, may be interpreted as the relative effect of that variable on the reading score for the population after controlling for all other variables in the model.

This analysis revealed that, with only one exception (New Brunswick French), the full model coefficients were not significantly different from the population model coefficients. The general conclusion is therefore that the set of variables used in the model do not account broadly for population differences. Nevertheless, intermediate stages of the model revealed a few effects of interest that are specific to New Brunswick French.

- Controlling for student demographics (gender, grade, born outside of Canada, language match, and number of books in the home) reduces the difference between New Brunswick French and the reference group, Ontario English.
- The greatest contrast between New Brunswick French and Ontario English is for number of books in the home. Ontario English students have, on average, many more books than New Brunswick French students.
- The coefficient for New Brunswick French increases as school characteristics, teacher characteristics, and student reading strategies are added to the model. This suggests that some of the variables in this category are suppressing performance of this population.

A few statistically significant changes for other populations are found in the intermediate models. These do not show up in the full model because of various complex interaction effects that were not analyzed. Examples are:

- For Saskatchewan English, the coefficient increases significantly when school characteristics are added to the model. The most likely source of this shift is that adding

this cluster controls for the relatively large number of Saskatchewan English schools with a high proportion of Aboriginal students.

- School characteristics also show a significant effect for Ontario English. Since this shift is relative to Quebec French (because Ontario English cannot be its own reference group), the most likely source of the shift is that this is a consequence of controlling for the large positive private school effect for Quebec French.
- For Quebec French, adding school characteristics significantly reduces the coefficient. The most obvious source of this shift is again the large private school effect for this population, compared to that for Ontario English. However, other positive school characteristics, including school size and class size, are also favourable to the Quebec French population.

The equity issue relates not to differences across populations but to variations in achievement within a population. Two equity measures were used: the total variance for each population (and its student and school components) and the interquartile range (the range between the 25th and 75th percentiles on the reading scale).

A plot of equity versus achievement confirms that francophone populations show less equity than anglophone populations but also shows that there is essentially no relationship between average achievement and equity.

Variables showing consistent differences between the two language groups include socioeconomic status, school and community size, class size, weekly class time on language arts, homework time, and student absenteeism. In all cases, these exert a stronger effect on francophone than on anglophone students. The greater variability (or less equity) found in francophone populations thus seems to be related to the differential effects of these variables on achievement for the two language groups.

The main conclusion from the study is that, although many of the variables available from the PCAP reading assessment contribute to achievement, few can be used as explanatory factors for differences in achievement across populations. This may be because differences between populations are much smaller than those between students and schools within populations. Much of the overall variation is thus masked by analytical techniques that focus on populations. Also, Canadian populations are actually more alike than different on many of the variables examined. Since most of these variables have only small effects on achievement in any case, most population-level effects are small relative to the statistical errors associated with these effects, and are thus difficult to detect.

Another possibility is that assessments such as PCAP are not optimally designed to detect the causes of achievement differences. In short, in emphasizing student, teacher, and school differences on the questionnaires, we may be measuring the wrong variables for examining population differences. For example, major reforms in some jurisdictions have not been considered, even though this seems to be useful in explaining international differences. Similarly, factors such as curriculum content and curriculum implementation (i.e., the content taught) are missing from the databases. Other societal factors such as expectations (achievement press) may also be important.

Finally, it may be unreasonable to expect that the measures of teaching and learning strategies captured by the PCAP questionnaires are adequately representative of students' overall schooling exposure. The effect of such factors, relative to demographic or socioeconomic factors, may simply be underestimated in the design of single-year, large-scale surveys.

TABLE OF CONTENTS

LIST OF TABLES AND CHARTS

6 Achievement Equity

Appendix

1 INTRODUCTION

Purpose

This report addresses the issue of jurisdictional[3] differences in school achievement in Canada and the related issue of equity in achievement. The key policy question being addressed is "how can Canadian educational jurisdictions attain the highest possible achievement levels combined with the least disparity between the highest and the lowest achieving students or groups?" The more specific analytical question is "what factors can help explain differences in achievement across Canadian jurisdictions?"

The primary data source for the report is the PCAP-13 2007 database. Other sources, particularly earlier Programme for International Student Assessment (PISA) and School Achievement Indicators Program (SAIP) reports, are used to determine if patterns exist for achievement levels and equity, and to identify characteristics of education systems in Canada that might be related to these outcomes.

Even a cursory review of the history of large-scale assessment in Canada reveals that differences between jurisdictions are both significant and persistent. There are also significant differences between the two official-language groups, both nationally and within jurisdictions. Factors related to achievement are also now being identified from research based on national and international databases. However, the relationship of these factors to jurisdictional differences has not been extensively investigated. Achievement equity has been even less widely studied, even though this is also of considerable policy interest because the school system is generally considered as a vehicle for bringing about greater societal equity.

Although often framed in socioeconomic terms, the equity issue goes beyond the effect of socioeconomic factors. In particular, there is increasing evidence that boys are doing less well in school than girls, particularly in reading. Rural-urban, ethnic, and other demographic differences are also of interest, as the policy goal is to reduce such disparities where possible. More generally, equity may be thought of as decreasing the difference between the lowest and highest achieving students. To the extent that greater equity can be achieved by increasing the performance of those at the lowest levels, this also contributes to increased average performance.

The literature review is intended to identify important patterns over time and place, while the PCAP database analysis will help corroborate these patterns or identify where specific features of PCAP yield either new results or results contrary to established patterns. The analytical phase will use several descriptive/comparative and regression analysis techniques for jurisdictional profiling and modelling of factors influencing achievement and equity. A combination of methods is needed because the "ideal" model for such work, namely a three-level hierarchical regression model, cannot be used because of the small number of units of analysis available at the jurisdictional level.

[3] In this report "jurisdictions" refer to provinces and territories. For much of the report, the main unit of analysis also incorporates the two official-language groups within jurisdictions where the minority-language group is large enough to permit analysis. This combination of jurisdiction and language is referred to as a "population" because the PCAP sample is explicitly stratified by these groups.

Research Questions

The research questions to be addressed in this report are:

a) What does the research literature, particularly that based on large-scale assessments, tell us about factors that contribute to jurisdictional differences in achievement?

b) Do some jurisdictions attain greater equity (lower variation) than others in achievement?

c) Were some jurisdictions able to narrow the gap between sub-populations (e.g., boys/girls, francophones/anglophones)? Is this related to or independent of average achievement?

d) What combinations of factors contribute to particularly high or low achievement within and across jurisdictions?

e) Are there particular combinations of school, teacher, and/or student characteristics that may offset the well-established effects of socioeconomic status on achievement?

f) Beyond the factors identified in the PCAP data, are there other characteristics of school systems in different jurisdictions that may contribute to differences in achievement levels?

g) Other than structural features, are there variables not being captured by large-scale assessments that may have a significant influence on achievement?

Conceptual Framework

The overarching conceptual framework for this report is an educational productivity model set within a human capital theory framework. That theory holds that the economic and social well-being of a society is determined not only by traditional stores of labour and capital but also by the knowledge and skills acquired by the population. This theory is the foundation of what is commonly called "the knowledge economy." A paper by Crocker (2007) gives details of this model. Its application to studies such as this is described in other reports (OECD, 2010; CMEC, 2011).

The value of education within the human capital model is typically captured by the "human capital earnings function" (Mincer, 1974; Becker, 1993). This function treats income (or other valued economic or social goods) as the outcome or dependent variable with education and experience (usually along with other covariates) as predictors or independent variables. The model is usually expressed as a regression equation, which can be used to evaluate the absolute and relative effects of specific predictors on the outcome.

The education production function is of the same form, but with achievement as the outcome and various inputs (e.g., resources, demographics) and processes (e.g., teaching and learning strategies) as independent variables. Contextual variables relevant to the system being analyzed are also sometimes used. Thus:

achievement = *f* (context, inputs, processes)

where "f" is a mathematical function, the form of which depends on the specific theoretical model and the type of analysis being performed.

Under this model, increasing average achievement and reducing disparities in achievement are treated as the immediate policy goals. Achievement, in turn, is assumed to have an impact on educational attainment, which in turn influences income and other longer-term outcomes within the broader human capital productivity model.

A number of attempts have been made since about the 1960s to develop conceptual models of school learning. Among the best known are the Carroll time model (Carroll, 1963, 1989) and the Wang/Haertel/Walberg (WHW) proximity model (Wang, Haertel, & Walberg, 1993).

The Carroll model is particularly interesting from a policy perspective because it captures a fundamental principle underlying the organization of school systems, that of the formal allocation of time to learning. This principle is manifested in macro-level policies such as compulsory school attendance and statutory lengths for school years and school days, mid-level policies such as time on particular subjects, and micro-level activities such as time on task in the classroom or homework assignments.

The WHW model of school learning advances a concept of "proximity" as a way of thinking about the relative effects of various factors influencing learning. The general hypothesis is that "proximal" factors, those that touch most closely on the day-to-day lives of students, are likely to be more influential than more "distal" factors such as administrative characteristics of the education system at the national level.

Some more recent syntheses are consistent with the WHW model and have also helped identify more specific positive influences on achievement. For example, Scheerens and Bosker (1997) produced a ranking of school factors that have positive influences on learning — their list included time, monitoring, pressure to achieve, parental involvement, and content coverage. Marzano (2003) independently developed a list that is almost identical to that of Scheerens and Bosker.

While these models help simplify thinking about factors affecting learning, the reality of school learning is that a large number of factors influence the desired outcomes. Existing theoretical frameworks, drawn mainly from psychology, sociology, and economics, can account for only a small number of such factors. Research based on large-scale assessments consistently shows that most measurable factors individually make only small contributions to learning. However, their cumulative effect is not well known.

Although none of these models is explicitly concerned with achievement differences, they are well suited to conceptualizing these differences because the logic of such models implies that cross-jurisdictional differences in the factors influencing achievement should also help explain achievement differences. In the following sections, we examine some of the factors that have been extensively studied in large-scale assessments, with emphasis on studies that focus explicitly on cross-jurisdictional differences.

Populations

The two official-language groups within jurisdictions are treated in this report as a jurisdictional variable for those jurisdictions with sufficient numbers in the minority-language group to warrant analysis. This is because the language groups function largely as independent systems. The jurisdiction/language combination is referred to as a "population" in PCAP, because the PCAP samples are explicitly selected from these groups. Where this is not the case, the minority-language group is combined across jurisdictions. For reasons discussed in the *PCAP-13 2007 Report on Differences in Reading Performances of 13-year-olds Based on Language and Minority/Majority Status* (CMEC, 2010), separate French-language groupings were formed for three eastern (Nova Scotia, Newfoundland and Labrador, and Prince Edward Island) and four western jurisdictions (British Columbia, Alberta, and Saskatchewan plus Yukon). Table 1.1 shows the populations thus formed, their sample and approximate population sizes for the PCAP-13 2007 reading assessment, and the mean and standard deviation for the reading component of that assessment for each population.

TABLE 1.1 **Jurisdiction/Language (Population) Groupings and PCAP-13 2007 Reading Scores**

Jurisdiction/Language	Abbreviation	Sample	Population	Mean	Standard Deviation
English					
British Columbia	BCe	1,646	51,647	486	90.50
Alberta	ABe	1,743	41,251	491	89.59
Saskatchewan	SKe	1,619	12,888	471	84.14
Manitoba	MBe	1,539	13,591	476	92.76
Ontario	ONe	1,651	157,085	503	93.98
Quebec	QCe	1,019	10,792	479	97.69
New Brunswick	NBe	1,540	6,400	466	89.53
Nova Scotia	NSe	1,611	10,999	471	90.89
Prince Edward Island	PEe	1,137	2,122	459	105.22
Newfoundland and Labrador	NLe	1,325	5,699	464	100.61
Yukon	YKe	179	395	486	93.07
French					
Quebec	QCf	1,179	85,831	532	111.79
Ontario	ONf	1,418	6,486	478	112.45
New Brunswick	NBf	1,467	2,575	458	112.37
Manitoba	MBf	474	1,471	436	111.73
West	Westf	329	591	471	112.68
East	Eastf	218	400	477	94.53
Total		**20,094**	**410,222**	**500**	**99.50**

Analytical Approach

At its simplest, the analytical model with achievement as outcomes and various contextual, input, and process variables as predictors should be applicable at the jurisdictional level. In this case, the unit of analysis is the jurisdiction, and the variables of interest are aggregated to the jurisdictional level (e.g., the outcome variable is average achievement for the jurisdiction rather than individual achievement).

Unfortunately, this simple model is not particularly useful because of the small number of units of analysis available (e.g., 13 provinces/territories or 17–20 units if the two official-language groups are treated separately). This means that any computed statistics will have large standard errors, making it difficult to detect statistically significant effects. Also, aggregating data to the jurisdictional level masks most of the variation. For example, in typical large-scale assessments, differences between students accounts for about 80% of the variation in scores, and differences between schools about 15%, leaving only about 5% of the total variation between jurisdictions.

Because of the multi-stage nature of the samples in large-scale assessments, it has become common practice to use hierarchical modelling to analyze the effects of student, teacher, and school factors on achievement. Typically, a two-level model, encompassing students and schools, has been used. The logical extension of this approach to the analysis of jurisdictional differences would be a three-level model, with jurisdictions as the third level. Work just completed for the *PCAP-13 2007 Factors Contributing to Performance in Mathematics and Science* report (CMEC, 2011) has revealed that it is technically feasible to use such a model. However, the results cannot be interpreted in terms of jurisdictional differences because the above issues of large standard errors and relatively small variation at the jurisdictional level are not solved by using a hierarchical model.

The approach to be taken here places less emphasis on mathematical modelling and more on jurisdictional profiling. We attempt to approach both jurisdictional differences and equity from several less formal perspectives. The general approach is as follows:

- Report in some detail the pattern of jurisdictional differences over time.
- Extract from the literature and from the PCAP 2007 data a set of factors that:
    - show consistent differences across jurisdictions;
    - are related (positively or negatively) to achievement;
    - are of policy interest, in the sense of being amenable to change through broad policy decisions that can be made at the jurisdictional level.
- Develop jurisdictional profiles in graphical terms, which depict the placement of jurisdictions on dimensions corresponding to the selected factors and the achievement scale.
- Examine more closely the variation in achievement levels within jurisdictions and examine the influence of selected factors on how this varies across jurisdictions (i.e., cross-jurisdictional differences in the spread of achievement scores).

Following this, an attempt is made to model both achievement and equity using two-level (student/school) models in which jurisdictions are entered as variables (i.e., each jurisdiction is coded 0 or 1) and the changes in coefficients for jurisdictions examined as other factors are entered into the model. In addition, separate analyses were conducted by jurisdiction and language accompanied by exploratory analysis of some specific predictor variables. The latter two approaches have been shown to be useful alternatives to three-level modelling for PCAP-13 2007 mathematics and science. These analyses are extended to PCAP-13 2007 reading in this report.

2 LITERATURE REVIEW

This chapter presents a brief review of literature on factors influencing academic achievement, with specific reference to the effects of these factors on interjurisdictional variation. The focus here is on international comparative research, especially that which attempts to account for differences among countries in educational achievement. The review draws heavily on a study by Haahr, Nielsen, Hansen and Jakobsen (2005), which examined country differences based on the PISA, PIRLS, and TIMSS databases and on reports of the 2009 PISA assessment, as these are considered to be the most comprehensive studies available on country differences. It also draws on the work of Hanushek, Wöessmann, and others who have examined resource and other factors in an international comparative context.

International Comparative Studies

Reporting on between-country differences in educational achievement has been the main goal of large-scale international assessments including the Programme for International Student Assessment (PISA), the Trends in International Mathematics and Science Study (TIMSS), and the Progress in International Reading Literacy Study (PIRLS). PISA is the most comprehensive of these studies, with assessments of reading, mathematics, and science, and occasionally other areas such as problem-solving and computer skills, being conducted every three years since 2000. Participation in PISA has grown from 32 countries in 2000 to 74 in 2009.

In all of these studies, the comparative "league table" results are widely reported. Differences between countries are relatively large, especially at the low end. For example, the lowest-performing country in PISA 2009 reading had an average score close to two standard deviation units (each SD unit is 100 points) below the OECD average of 500. Several countries had average scores more than one standard deviation unit below the OECD average. At the high end, the scores are less extreme, with the highest-performing country at about one-half standard deviation above the OECD average and a large number of countries clustered close to that average (OECD, 2010a).

Canada has historically been near the top of the distribution of PISA scores. Canada has also sampled in sufficient detail to allow provincial results to be reported. Provincial differences have averaged close to one-half a standard deviation unit (50 points), with the highest-performing provinces being close to the top of the international rankings and the lowest-performing provinces close to the average[4] (Bussière, et al., 2001, 2004, 2007; Knighton, Brochu, & Gluszynski, 2010).

Some inferences about country differences can be made from straightforward examination of the results. For example, it is obvious that most of the lowest-performing countries are from the less developed parts of the world. However, performance also varies quite widely among the developed countries that make up the membership in OECD. There has also

[4] The performance of Canada and the provinces on PISA is examined in more detail in the next chapter.

been much discussion and speculation on why some countries such as Finland or Korea have done so well and even on why two highly similar countries, such as Canada and the United States perform differently (Willms, 2004).

In addition to the league tables, a large number of research reports focusing on factors contributing to achievement have emerged from the large-scale assessment work. Nevertheless, relatively few studies have directly addressed the sources of inter-country differences or attempted to explain these differences.

The investigation of inter-country differences is plagued by methodological challenges, arising mainly out of the fact that differences between countries are much smaller than differences between students within countries. Aggregating data to the country level thus masks most of the differences and leaves few units of analysis with which to work (Jürges & Schneider, 2004; Fuchs & Wöessmann, 2004). Looking at simple correlational effects at the country level is not adequate both because of the small number of units and because other factors are not controlled.

Factors Influencing Achievement

A large number of factors have been found to influence achievement. Most of the effects are small and tend to be in the same direction (though of widely different magnitudes) in their influence on achievement across countries (e.g., OECD, 2010h). These effects have been reviewed in some detail in another report in this series (CMEC, 2011). In general, the observed effects have been consistent with the Carroll and Wang/Haertel/Walberg models. However, socioeconomic and demographic effects have tended to be stronger than the effects of teaching and learning strategies in these studies. This is likely an artifact of the measurement instruments used, in which the measured socioeconomic and demographic effects, as well as achievement scores, tend to be long-term and stable, whereas teaching and learning effects can be measured only for the year of the assessment.

The remainder of this chapter looks at variables under a number of major categories designed to capture both background factors (socioeconomic and demographic), resource effects, and teaching and learning effects. Rather than reviewing a large number of such factors in detail, the review focuses on the limited number of studies that have been more explicitly focused on accounting for interjurisdictional differences.

Socioeconomic Status

Socioeconomic status (SES) is important in discussing jurisdictional differences, especially in the international context, because SES differs widely across countries and because there is strong evidence that lower achievement is more prevalent among those from lower socioeconomic backgrounds. SES is also particularly important in examining the equity issue because countries differ widely in the distribution of SES levels and the magnitude of the SES effect on achievement.

Haahr et al. (2005) argue that although SES matters significantly for student academic performance, the degree to which socioeconomic background matters differs significantly across countries. In other words, some countries perform better than would be expected; others worse. The findings support the notion that countries that focus on equity of outcomes in policy decisions can compensate for different socioeconomic backgrounds.

> For some of the best performing countries in terms of average achievement scores (Finland, Hong Kong–China, Japan, and Korea), the socioeconomic background of students thus matters the least (Haahr et al., 2005, p. 10).

Haahr et al. (2005) also found that, in PISA 2003, for Finland, Japan, Italy, Latvia, and Iceland, the socioeconomic characteristics of students explain only about 10% or less of total variance in performance, whereas the corresponding figure is over 20% for Germany, Portugal, Belgium, and Hungary. The authors note that four of the top performing countries (Finland, Korea, Hong Kong–China, and Japan) are among the seven countries where the impact of student socioeconomic background is smallest. The overall conclusion is that since the relationship between socioeconomic background and performance varies between countries, then factors that differ between countries are affecting the relationship (Haahr et al., 2005, p. 99).

Willms (2003, 2006) and others have argued that a desired goal of schooling is to reduce the "socioeconomic gradient" or the degree of inequality across socioeconomic groups. Willms specifically examined the relationship between reading performance and SES. Key findings are that the analysis of socioeconomic gradients can help to provide direction to policy developers from national, provincial, school district, or school-based level depending on specific findings.

Among the PISA 2009 reports is a special volume on issues pertaining to achievement equity (OECD, 2010b). Among the high-performing countries and economies, Shanghai–China, Korea, Finland, Hong Kong–China, and Canada all show gentle slopes of the socioeconomic gradients, suggesting that large differences in student socioeconomic backgrounds are, on average, not associated with large performance differences among students. Among countries and economies in which students perform slightly below or around the OECD average, the same is true for Estonia, Iceland, Portugal, Italy, Spain, and the partner countries and economies Liechtenstein, Macao–China, and Latvia.

In these countries, a relatively smaller proportion of low-performing students come from disadvantaged backgrounds and the relationship between school performance and schools socioeconomic intake is weaker. Thus, by themselves, policies that specifically target students from disadvantaged backgrounds would not address the needs of many of the country's lower performing students. Moreover, policies targeted as SES in these countries would be providing services to a sizeable proportion of students who already perform well (OECD, 2010b, p. 108).

Demographic Factors

Student-level factors usually investigated under this heading include gender, race, ethnic or immigrant background, and home language relative to the language of the school. These are sometimes also aggregated to the school level and included with school or community characteristics such as school size, class size, the ethnic or racial mix found in the school, school configuration (e.g., grade levels served), or community size.

Among the strongest of these findings is that for achievement by gender and language. Almost all large-scale studies show a strong effect in favour of girls in reading assessments (Knighton et al., 2010; CMEC, 2009). In contrast, the same studies show smaller and less consistent results for mathematics and science.

Haahr et al. (2005) report that in all countries females achieve significantly higher average scores in reading than males, with the score point difference ranging between 21 and 58 score points. These authors concluded that there is potential gain in increasing the focus on the reading performance of boys, especially in countries where the differences are largest.

Language as well as racial and ethnic effects tends to be country-specific because of different language configurations and immigration patterns. Haahr et al. (2005) reported that in schools where at least 40% of students' first language is not the language of the country concerned, non-native speakers have significantly lower achievement than in schools with low density (10% or less) of non-native speakers. Again, this can influence the overall performance of students in countries with large numbers of immigrants. Indeed, some of the highest-performing countries in PISA are those who not only have high SES but also relatively homogeneous populations.

In Canada, the SAIP, PISA, and PCAP assessments show that anglophone students have higher achievement scores than francophone students, except in Quebec, where both language groups tend to do equally well. This suggests that francophone students do not perform as well when in a minority-language situation. This affects the overall performance of those few jurisdictions with high proportions of minority francophone students, as well as differences between the majority and minority groups in these jurisdictions, another issue of equity.

Resource Effects

Research on school effects has most often emphasized the allocation of resources to schools and their use. This can be extended to jurisdictions (whether provinces, states, or countries) because most resource-allocation policies are made at the jurisdictional level. Indeed, many resource studies, including much of Hanushek's work, are based on state- or country-level comparisons of resource allocations.

Resource allocations are commonly measured by such indicators as per-student expenditures or by variables such as teacher qualifications or class sizes, which are directly

related to cost. Evidence about how resources are used is less common. It is not possible to do justice to this complex area in this review. However, a few examples can be given of results that seem to bear most directly on jurisdictional differences.

Hanushek has consistently argued that the research shows no strong or consistent evidence to support a relationship between educational resources and student achievement. This conclusion is based primarily on comparative analyses of state-level differences in resource allocations in the United States or country-level differences in international studies. Hanushek and Wöessmann support the view that some uses of resources can be effective, but argue that much of the increase in resources goes into activities that have little or no effect or have shown "inconsistent outcomes" (Hanushek and Wöessmann, 2010, p. 35).

A RAND Corporation review and meta-analysis (Grissmer, Flanagan, Kawata, & Williamson, 2000) examined US policies that encourage higher allocation of resources to disadvantaged groups. The hypothesis was that these resources should result in disproportionate gains for the targeted groups. The analysis supported this hypothesis, with the largest gains in National Assessment of Educational Progress (NAEP) scores being found for Blacks, Hispanics, and lower-achieving white students. The authors interpreted this result as indicating that such policies have the desired payoff in equity terms, although this seems to have had little impact on overall average performance. In terms of jurisdictional differences, these results suggest that policies designed to improve the performance of lower-achieving students are likely to have a greater effect on equity than on overall average. This is especially true if the proportion of students receiving greater resources is small.

Class size deserves comment because of its high public profile and the strong common-sense belief that smaller classes should yield higher achievement (and other benefits). However, significant reductions in class size can come only at considerable cost because of the additional teachers and space required. Reducing class size is an explicit policy direction in some jurisdictions, suggesting that this is an important variable to examine in accounting for jurisdictional differences.

Experimental studies of class size in the early grades (Finn & Achilles, 1999; Molnar et al., 1999; Nye, Hedges, & Konstantopoulos, 1999) seem to show that smaller classes have positive effects on achievement in the lower grades. However, the results from large-scale assessments, which typically encompass students in middle or secondary grades, tend to show the opposite class size effect (those in larger classes have higher scores) even when other school factors are controlled (CMEC, 2003, 2005, 2008).

It is important to note that these assessments do not account for the possibility that lower-achieving students may be assigned to smaller classes, thus creating an artificial effect. Nevertheless, these studies do suggest that class size is certainly no stronger than many other factors as a determinant of achievement, especially in the intermediate and secondary school grades. It follows that jurisdictional differences in class size policies are not likely be a useful factor in accounting for jurisdictional differences in achievement.

Other School-Level Effects

Aside from resources, other school-level variables that have been investigated include school size, school governance, community size and location, school decision-making and autonomy, parent engagement in the school, and some student-level variables such as socioeconomic status aggregated to the school.

The relative performance of public and private schools has been the subject of considerable investigation. A review of international studies (Coulson, 2009) concluded that private sector schools outperform public sector schools in the overwhelming majority of cases, and that the results are more pronounced in situations where "private" schools are driven by market forces such as competition.

Similar public/private differences are found in most of the large-scale assessment studies (OECD, 2001, 2004). This is an obvious issue of interjurisdictional difference because the level of acceptance of private schools or, more generally, choice in schooling is a matter of jurisdictional policy.

Similar results were found in a comparative OECD study based on the PISA 2003 assessment (Wöessmann, Lüdemann, Schütz, & West, 2007). Students perform better in countries with more choice and competition as measured by the share of privately managed schools, the share of total school funding from government sources, and the equality of government funding between public and private schools.

The issue of school autonomy and locus of decision-making has also been the focus of many studies of school effects. The Wöessmann et al. study cited above is an example. That study and more recent work by Hanushek and Wöessmann (2010) concluded that different facets of accountability, autonomy, and choice are strongly associated with the level of student achievement across countries. Students perform better where policies are in place that focus on students (external exit exams), teachers (monitoring of lessons), and schools (assessment-based comparisons).

Haahr et al. (2005) found that student performance is positively correlated to the percentage of schools that have responsibility/autonomy for teacher appointments, disciplinary policies for students, teacher dismissal, course content, and course offerings. Hanushek and Wöessmann (2010) in their discussion of institutional structures and incentives in the school system, report similar findings.

Students in schools with hiring autonomy perform better on average, while they perform worse in schools that have autonomy in formulating their budget. School autonomy over the budget, salaries, and content of courses appears to be more beneficial when external exit exams hold schools accountable for their decisions. Similar results were found in a study by Gunnarsson, Orazem, Sanchez, and Verdisco (2009), based on data from 10 Latin American countries. However, the autonomy effect disappeared once the factor of choice to exert autonomy was controlled. The authors conclude from this that autonomy cannot be imposed by central authorities — it must be chosen by school administrators.

A clear statistical correlation exists between the degree of institutional differentiation (i.e., tracking or streaming students) and the variance of student performance (Haahr et al., 2005). In other words, dividing students into tracks increases disparity on the one hand, and does not improve average performance on the other; weak students become even weaker; strong students, even stronger.

Institutional differentiation also means that socioeconomic background matters more; the notion here is that teachers become less, not more, likely to support students in countries where there is a great deal of student streaming and tracking (Haahr et al., 2005).

Haahr et al. (2005) argue that educational spending per student is not a determinant of PISA or TIMSS country student outcomes, particularly in light of comparison of outcomes of affluent countries. In other words, even though there is a positive relationship between the Gross Domestic Product (GDP) per capita and mean achievement scores, some countries perform above expected outcomes predicted from their GDP per capita; other countries perform worse than would be predicted.

These authors also conclude that policy reform can be effective. There is scope for improvement within given economic resources. This is illustrated by the cases of Poland and Latvia, which have made reforms that have resulted in significant improvements in average scores in PISA tests.

> Notably, improvements in the Latvian education system have taken place without a greater share of available economic resources being devoted to education. From 1998 to 2002, public expenditure on education as a percentage of GDP fell from 6.29% to 5.82%, albeit in the context of relatively strong economic growth. In the same period, the share of Polish public expenditure devoted to education grew from 5.09% of GDP to 5.6% of GDP, an increase of 10% (Eurostat, 2005) (Haahr et al., 2005 pp. 68-69).

Finland is also cited as standing out as a high achieving country that has brought in education reforms based on the principle of equity and of minimizing low achievement.

> A comprehensive network of schools and the recruitment of highly qualified teachers in all schools have been important means in ensuring educational equality in all regions of Finland (Haahr et al., 2005, p. 71).

Finally, these same authors also point out that Finland has been successful in achieving high PISA scores as well as very low between-school variance in scores. However, other countries, for example, Denmark have both high achievement in PISA mathematics scores and high between-school variance. The authors conclude that similar student performance across schools is compatible with, but not a necessary precondition for, high performance standards.

Instructional Climate

Components of instructional climate include time allocation and use, homework, and absenteeism, as well as assessment practices and the accommodation of special-needs students.

Haahr et al. (2005) found large differences among countries in time allocated to learning as measured by average length of periods, number of classes per week, and learning outside the classroom. According to the authors, this issue is not clear-cut from the data. For example, students from Korea, a high-performing country, reported almost 50 hours of study per week, while those from Finland, Netherlands, and Japan, three other high-performing countries, reported around 30 hours of study per week.

Homework is one important component of time that can vary substantially across students, schools, and countries. A comprehensive review of the effects of homework on academic achievement is available (Cooper, Robinson, & Patall, 2006). The results show the effects of homework to be generally positive. Effects are very small at the elementary level but increase at the higher grades. Again, it is not completely clear whether these results simply reflect the fact that better students do more homework. However, as a teaching strategy, it is more plausible to encourage homework than to simply assume that time on homework follows ability.

According to Haahr et al. (2005), data from PISA, TIMSS, and PIRLS indicate that there is a weak statistical relation between time reported to be devoted to learning and achievement. Weaker students may spend more time doing the same homework as their higher achieving counterparts or may even be assigned more homework. The authors contend that resources may be better spent on improving the teaching and learning experience. Other researchers (e.g., Fuchs and Wöessmann, 2004) found a positive correlation for homework and achievement for mathematics and science but not for reading.

Most studies of school absenteeism are concerned with factors contributing to absenteeism rather than to the impact of absence on achievement. Days absent showed negative effects on mathematics achievement in the SAIP mathematics assessment (CMEC, 2003) and on reading achievement in PCAP-13 2007 (CMEC, 2009). In reality, most students in Canada attend school regularly, so it is difficult to examine the effect of high absenteeism rates. Absenteeism seems not to have been investigated in the international studies.

School assessment practices have been investigated in some studies as a factor contributing to jurisdictional differences. In particular, the presence or absence of external exit examinations is an important differentiating feature of national education systems and sometimes of jurisdictions (states or provinces) within systems. Jürges and Schneider (2004), Haahr et al. (2005), and Hanushek and Wöessmann (2010) have provided evidence that external exit exams have positive effects on achievement. In particular, Hanushek and Wöessmann (2010), citing studies by Bishop (1997) and Jürges, Schneider, & Büchel (2005), note that

> In Canada and Germany, the two national education systems where the existence of external exams varies within the country because some regions feature them and others do not, it has similarly been shown that students perform better in regions with external exams (Hanushek & Wöessmann, 2010, p. 38).

More general measures of school climate are derived from ratings of principals on a number of aspects of climate, which may be aggregated to form generic school climate or instructional climate variables. Haahr et al. (2005) report that positive school climate accounts for less than 4% of total variance in PISA 2003. Nevertheless, the top seven countries where principals were most positive about school climate included four out of the top six performing countries across all disciplines. The authors link school climate to school differentiation or streaming. Specifically, three of the four countries where school climate factors matter the most are also countries where institutional differentiation is most extensive.

One of the volumes based on the PISA 2009 data (OECD, 2010d) addresses the question: "What makes a successful school?" in the context of country differences. The report points to Poland, Korea, and other countries, which have made significant improvements in equity, and have thereby raised the average performance of their students. The obvious conclusion from this is that focusing on equity does not have to come at the expense of high-performing students. Other major findings of that study are:

- Equity of learning opportunities creates an environment that minimizes the impact of SES. On the other hand, school systems that have sought to deal with student needs through differentiation have not produced superior results and in some cases have lower results.
- At the country level, greater individual school autonomy appears to be a key aspect of success especially in terms of what is taught, and how assessment occurs.
- Within countries, autonomy of resource allocations and public reporting of results, which lend a degree of accountability, go hand in hand.
- Higher teacher pay rather than smaller classrooms produces better student performance at the level of the school system.
- While highly valuing education and providing education resources are important, nevertheless, within a given range, increasing the level of resources does not affect outcomes.
- Attendance at pre-primary education has a positive effect on student performance.
- Reading levels are positively affected by a strong disciplinary climate, positive teacher behaviour, and good teacher-student relations.

All of these findings are consistent with other studies and suggest specific reforms that could occur at the system level, which could improve performance and thereby reduce jurisdictional differences.

Human Capital Research

Hanushek and Wöessmann (2010) combined data from international tests given over 45 years to develop a comparable index of skills of people in the labour force. This is then used to examine the policy question of whether it is better to pursue basic skills as opposed to elite-level skills as it relates to growth in OECD countries. From this analysis of micro-level data on international assessments, they conclude that:

> While small samples and collinearity obviously suggest caution in the interpretation of these results, it seems clear that basic skills are relevant for OECD-country growth. Furthermore, if anything, high-level skills appear more important in non-OECD than in OECD countries, rather than the other way around (Hanushek & Wöessmann, 2010, p. 18).

Their analysis also shows that we can use proxy measure of cognitive skills on international achievement tests as a way to understand differences in growth in OECD countries over time. While they acknowledge that the estimated impact on growth will vary depending on the theoretical framework chosen, the authors' analysis shows that the impact of education polices remain significant and "yield enormous values no matter what" (Hanushek & Wöessmann, 2010, p. 40).

Having made the argument about the importance of basic skills, these authors examine the determinants of educational achievement. A summary of their conclusions follows.

- More resources (i.e., class size/more funds for schools) do not make a difference.
- Teacher quality is important in improving student achievement. However, differences in teacher quality are related to teacher behaviours, and these do not closely correlate with the usual measures of teacher quality (i.e., teacher education and experience).
- Acknowledging the challenge in measuring teacher quality and in regulating it as well, the authors suggest that the research may point the way to policy approaches that encompass competition among schools, accountability and student testing, and autonomy of schools at the local level.
- In terms of economic growth in highly developed countries, achievement of basic skills has more payoff than achievement of high skills and tertiary schooling in the long run.

Canadian Research

The only previous Canadian study known to address explicitly jurisdictional differences is that by Ma and Crocker (2007). That study used the PISA 2000 Canada database to examine factors related to differences between provinces. Given the difficulties noted earlier in the use of three-level hierarchical models, that study took the approach of using two, two-level models, the first with students nested within schools and the second with schools nested within provinces. The residual school means from the first model were used

as input to the second model. This was followed by graphical profiles designed to show the differential effects of variables on student performance.

That study found significant differential effects across provinces for a number of school context and climate variables, as well as student and teacher behaviours. Because of the Canadian context and because of the close connection to this study, these effects are summarized in Table 2.1.

Focusing on differential jurisdictional effects of predictor variables on achievement, particularly if some are positive and some negative, highlights the potential for "local" effects of some variables. However, this does not help account for actual differences in achievement scores. The reality is that most of the factors influencing achievement have similar effects, at least in direction, across jurisdictions. The key to understanding jurisdictional differences in achievement lies not so much in finding the differential effects of the predictors as in the difference in values of these predictors across jurisdictions. If a jurisdiction can be found to have a particularly high or low value for some variable related to achievement, and if that jurisdiction also has particularly high or low achievement, then that predictor is of interest in an explanatory model. The problem is how to control for the effects of other predictors, when normal modelling techniques break down.

TABLE 2.1 **Differential Effects on Reading Performance Across Provinces for Selected PISA 2000 Variables**

Variable	Effect
Student/teacher ratio	Positive for ON Close to zero for other provinces
School location (community size)	Positive for BC Negative for AB and MB Close to zero for others
School material resources	Negative for QC Close to zero for others
School instructional resources	Negative for QC and ON Close to zero for others
Disciplinary climate	Positive for all Stronger for ON and BC than for others
Achievement press	Negative for NL, QC, MB, BC Close to zero for others
Student behaviour	Positive for most Negative for NB
Sense of belonging to school	Positive for most Close to zero for NS, NB
Teacher morale	Strongly positive for Ontario Slightly positive for most others Slightly negative for BC
Instructional time on language arts	Mostly positive Close to zero for some

(Source: Ma & Crocker, 2007)

In summary

The most significant problem in reviewing research on jurisdictional differences is that of distinguishing between direct influences on achievement and those influences that can help account for differential achievement across jurisdictions. This requires both that a particular variable have an effect on achievement at the individual student (or school) level and also that it vary across jurisdictions sufficiently to influence average achievement at the jurisdictional level. For reasons given earlier, there is no simple way to model jurisdictional differences. Most studies rely either on simple correlational relationships using data aggregated to the jurisdictional level or on models that are inherently limited in their ability to examine complex patterns at the jurisdictional level.

Nevertheless, the research on jurisdictional differences does point to a number of important conclusions. These may be summarized as follows:

- Many factors influence school achievement. None has a decisive effect, and the cumulative effects of all factors included in the various studies have not been well established.
- Differences within jurisdictions are more important than differences between jurisdictions.
- Broad policies and reforms at the jurisdictional level can be important, as evidenced by improvements in achievement in some countries that have instituted major reforms.
- Absolute resource levels seem not to matter as much as how the resources are used. Within a given range, increasing resource levels seems to make little difference to achievement.
- Socioeconomic status is important. Almost all low-achieving countries have relatively low average SES. However, not all high SES countries have high achievement. The effects of socioeconomic status can be mitigated by policies designed to promote equity.
- High achievement and a high degree of equity are not conflicting goals. Some countries, including Canada, have achieved both high performance levels and a relatively high degree of equity among students.
- Countries with highly differentiated schools, through tracking or streaming of students have variable achievement but less equity in achievement.
- Internationally, having a first language other than the language of the school tends to have a negative effect on achievement.
- Gender differences strongly favour girls in reading. The size of the gender gap varies widely across countries, suggesting that reading achievement can be improved in some jurisdictions by focusing on improving the performance of boys.
- There are indications that systems with more frequent testing and external exit exams tend to have higher achievement.

- The high level of both achievement and equity in Finland has been attributed to the existence of a comprehensive network of schools and the recruitment of highly qualified teachers in all schools.
- In terms of economic growth in highly developed countries, achievement of basic skills has more payoff than achievement of high skills and tertiary schooling in the long run.
- The overall performance of Canada has declined slightly over the past decade. Nevertheless, Canada remains one of the highest-performing countries while combining this with a high degree of equity.

3 TRENDS IN CANADIAN INTERJURISDICTIONAL DIFFERENCES

Jurisdictional Achievement Relative to the Canadian Average

Most of the reports have presented the comparative results in terms of whether a jurisdiction is lower, the same as, or higher than the Canadian average, based on mean scores and confidence intervals. This approach was taken as the starting point for summarizing the data.

For purposes of this report, the combination of jurisdictions and official-language groups within jurisdictions is used as the main unit of analysis. Where available, the comparative results are thus reported in terms of jurisdiction/language, or what is called "populations" in SAIP/PCAP terminology.

Table 3.1 gives the total number of achievement measures available for all populations for which numbers are sufficient, based on all assessments from 1993 to 2009.[5] These have been grouped to allow for some comparison over time as well as across jurisdictions. Specifically, the first two cycles of SAIP, from 1993 to 1999, the third cycle of SAIP, from 2001 to 2004, the three available PISA cycles, and the single PCAP cycle, are reported separately for this comparison. The number of measures is much larger than the number of assessments because each assessment has multiple measures. In this case, only the results at the subject level are reported. Results for sub-tests within the subjects closely follow the subject-level pattern.

Out of a total of 556 achievement measures in the table, approximately half gave results that were low compared to the Canadian average (CAN), and only 10% were high. This is partly a consequence of the weighting used to compute the Canadian average. While larger jurisdictions have higher weight, the lower scores tend to occur in the smaller jurisdictions. This has a particular impact on Ontario English, because that population contributes such a large amount to the Canadian average that its results are more likely than any other to be close to that average.

The jurisdictions are grouped in Table 3.1 based on a rough ranking developed by computing the percentage of "lows" for each jurisdiction and ordering these from the smallest to the largest number of lows. Within a province, English and French populations are kept together even though this is not consistent with their rankings. Using the number of "highs" yields much the same result. However, the ranking is less stable because there are many fewer highs than lows.

[5] The populations used in this chapter are drawn from the available reports and thus differ slightly from those in Table 1.1.

TABLE 3.1 **Number of Measures and Percentages Lower than, the Same as, and Higher than the Canadian Average**

	SAIP 1993-1999			SAIP 2001-2004			PISA 2000-2009*			PCAP-13 2007		
	Low	Same	High	Low	Same	High	Low	Same	High	Low	Same	High
Total measures	114	153	29	80	46	11	70	30	20	38	6	4
Jurisdictions	Percentage			Percentage			Percentage			Number of measures**		
QCf	0	56	44	0	83	17	25	67	8	0	0	3
QCe	0	78	22	17	67	17				2	1	0
AB	0	56	44	13	13	75	0	17	83	1	1	1
BC	17	83	0	11	89	0	0	25	75	3	0	0
ONe	44	56	0	13	88	0	17	83	0	0	3	0
ONf	100	0	0	75	25	0				3	0	0
MBe	28	67	6	38	63	0	67	33	0	3	0	0
MBf	50	50	0	63	13	25				3	0	0
YK	39	61	0	63	38	0				3	0	0
NSe	22	78	0	100	0	0	100	0	0	3	0	0
NSf	50	8	42	63	38	0				2	1	0
SK	25	67	8	75	25	0	75	25	0	3	0	0
NBe	33	67	0	88	13	0	100	0	0	3	0	0
NBf	67	33	0	63	25	13				3	0	0
NL	39	61	0	75	25	0	100	0	0	3	0	0
PE	33	56	11	100	0	0	100	0	0	3	0	0
NT	100	0	0	100	0	0						
NU***	100	0	0	100	0	0						
CAN	**39**	**52**	**10**	**58**	**34**	**8**	**58**	**25**	**17**	**79**	**13**	**8**

* Language breakdowns by province are not available for PISA in all years.
** The number of measures rather than the percentage is used for PCAP because of the small number of measures.
*** Nunavut results are combined with those for the Northwest Territories prior to 1999.

It is clear from these results that Alberta and Quebec are the highest-performing jurisdictions overall. Within Quebec, francophone students tend to do better than anglophone students, though the performance of both language groups is better than that for other jurisdictions. In general, the lowest performance is found in the Atlantic provinces, Saskatchewan, the Northwest Territories, and Nunavut. Ontario is more likely than others to be found near the Canadian average because of the high weight contributed by Ontario to that average.

The broad rankings are remarkably stable over time, especially for the higher and lower performing populations. The main anomaly in the pattern is the relatively high performance of British Columbia on the PISA assessments. There is no way to tell if this represents a significant shift over time or if the results are specific to the PISA measures. However, it is notable that other jurisdictions do not show a similar shift in their relative positions. It should also be noted that these results are relative and give no information on whether all populations are changing over time.

Magnitude of the Differences

While this type of ranking clearly points to the consistency of differences across a wide range of measures, it says nothing about the magnitude of these differences. Even small differences show up in rankings, although it would be unusual to find such a consistent pattern if the differences between the highest and lowest populations are small. It is thus useful to give some indication of the size of the observed differences, in terms of the measurement scales used and in terms of other difference indicators, such as differences between schools or students.

The SAIP results were reported as the proportion of students at specific levels on a five-point scale. These were also summarized more concisely as the proportion at or above an "acceptable" standard, defined as Level 2 for 13-year-olds and Level 3 for 16-year-olds. Table 3.2 shows these latter proportions on selected measures for the top three and bottom three jurisdictions. The territories have been omitted from these tabulations because the results for the territories, with the exception of Yukon, tend to be extremely low, even in comparison to the lowest figures given in the table.

These results confirm the previous pattern, with Quebec and Alberta dominating the top scores and the Atlantic provinces, along with French populations outside Quebec, appearing most often among the lowest scores. Overall, close to 30 percentage points separate the highest from the lowest scores in this table. While close to 80% of students in the top-ranked populations meet the minimum performance criterion, this is the case for only about 50% at the lowest levels.

TABLE 3.2 **Percentage of Students at Minimum Acceptable Level for Highest and Lowest Ranked Three Populations on Selected SAIP Measures**

Year	Subject	Age	Highest Rank			Lowest Rank		
			1	2	3	3	2	1
1993	Mathematics	13	80 QCf	69 QCe	67 AB	54 PE	53 NL	48 MBe
		16	74 QCf	64 MBf	63 QCe	52 MBe	48 PE	47 NL
1994	Reading and Writing	13	82 QCf	79 AB	79 QCe	74 MBe	74 NBe	66 NBf
		16	80 QCf	74 ABe	74 QCe	62 MBf	61 ONf	60 NBf
1996	Science	13	83 AB	76 PE	76 SK	60 NBf	60 MBf	57 ONf
		16	79 AB	73 QCf	71 SK	64 NL	58 NBf	51 ONf
1997	Mathematics	13	78 QCf	65 QCe	65 AB	52 MBe	50 ONe	48 SK
		16	81 QCf	74 QCe	63 NBf	49 PE	47 NBe	43 NL
1998	Reading and Writing	13	84 QCf	78 NL	78 AB	71 NSe	71 MBf	58 NSf
		16	79 QCf	72 QCe	72 ONe	64 PE	62 NSf	60 MBf
1999	Science	13	83 AB	76 SK	74 PE	61 MBf	61 NBf	57 ONf
		16	86 AB	81 PE	81 QCf	72 ONe	69 NBf	60 ONf
2001	Mathematics	13	75 QCf	71 AB	67 QCe	52 NBe	49 NSf	48 NSe
		16	63 MBf	61 AB	56 NSf	42 SK	42 ONf	36 NL
2002	Writing	13	88 QCf	85 ONe	83 MBe	75 NL	75 MBf	73 NSf
		16	75 QCf	67 QCe	60 MBe	45 ONf	43 MBf	43 NSf
2004	Science	13	78 AB	73 QCf	72 ONe	31 PE	31 ONf	29 NL
		16	72 AB	66 QCf	64 ONe	58 NBe	57 NBf	48 ONf
Average		13	81	75	73	59	58	54
		16	77	70	67	56	54	50

A similar pattern is evident from the PISA scores in Table 3.3. The main difference in provincial patterns is that Ontario and British Columbia appear more often among the top three than was the case for SAIP. In almost all cases, the Atlantic provinces represent the lowest three. The PISA numbers are remarkably stable across subjects and over time. Since the PISA scores are based on a scale with mean 500 and standard deviation 100, the difference between the highest- and lowest-performing provinces is close to one-half standard deviation. Placed in an international perspective, the highest-performing provinces are comparable to the highest-performing PISA countries, while the lowest ones are at approximately the OECD average.

TABLE 3.3 **Mean Scores for Highest and Lowest Ranked Three Provinces on PISA Assessments**

Year	Subject	Highest Rank			Lowest Rank		
		1	2	3	3	2	1
2000	**Reading**	550 AB	538 BC	536 QC	517 PE	517 NL	501 NB
	Mathematics	550 QC	547 AB	534 BC	512 PE	509 NL	506 NB
	Science	546 AB	541 QC	533 BC	526 NS, NL	508 PE	497 NB
2003	**Reading**	543 AB	535 BC	530 ON	512 SK	503 NB	495 PE
	Mathematics	549 AB	538 BC	537 QC	515 NS	512 NB	500 PE
	Science	539 AB	527 BC	520 QC	489 PE	498 NB	505 NS
2006	**Reading**	535 AB	534 ON	528 BC	505 NS	497 PE	497 NB
	Mathematics	540 QC	530 AB	526 ON	506 NS	506 NB	501 PE
	Science	550 AB	539 BC	537 ON	527 SK	509 PE	506 NB
2009	**Reading**	533 AB	531 ON	525 BC	499 NB	495 MB	486 PE
	Mathematics	543 QC	529 AB	526 ON	503 NL	501 MB	487 PE
	Science	545 AB	535 BC	531 ON	506 MB	501 NB	495 PE
Average		545	537	531	512	507	501

The PCAP figures are given in Table 3.4. PCAP is scored on the same scale as PISA. Overall, the pattern of highest- and lowest-performing groups is much the same as in the other assessments. The difference between the highest and lowest average scores is larger for PCAP than for PISA, at close to three-fourths of a standard deviation.

TABLE 3.4 **Mean Scores for Highest and Lowest Ranked Three Populations on PCAP-13 2007 and PCAP 2010 Assessments**

Year	Subject	Highest Rank			Lowest Rank		
		1	2	3	3	2	1
2007	**Reading**	532 QCf	505 ABf	491 ABe	459 PE	458 NBf	436 MBf
	Mathematics	518 QCf	510 QCe	508 ONe	460 NBf	457 NSe	449 PE
	Science	524 ABe	518 QCf	499 ONe	467 QCe	464 PE	460 NBf
	Average	525	511	499	462	459	448
2010	**Reading**	517 ONe	506 ABe	499 BCe	468 SKf	464 NBf	464 YKe
	Mathematics	516 QCf	511 ONf	507 ONe/NBf	503 NSf	498 SKf	480 MBf
	Science	515 ABe	510 ONe	506 ABf	482 MBf	482 NBf	478 YKe
	Average	516	509	504	484	481	474

Another way of looking at the magnitude of differences is through the relative amount of variation in scores between jurisdictions compared to that between students and schools. This is done only for PCAP-13 2007. Table 3.5 shows these differences, expressed as standard deviations (on the PCAP achievement scale, with a Canada mean of 500 and standard deviation 100) at the student, school, and population levels. This shows that most of the variation in achievement is at the student level, with less at each of the school and population levels.

TABLE 3.5 **Student, School, and Population Standard Deviations: PCAP-13 2007**

Subject	Standard deviation		
	Student level	School level	Population level
Reading	99.5	51.4	22.1
Mathematics	99.4	57.0	18.6
Science	97.1	62.3	16.7

The Stability of Results and the Persistence of Differences over Time

Taken at face value, these results suggest that there has been little change in the overall performance of Canadian students on these assessments over the time under review. We hesitate to make a strong statement to this effect because there is insufficient evidence on whether the measures themselves have been stable over time. However, the SAIP assessments were explicitly criterion-referenced, with levels of acceptable performance established by expert panels external to the test developers. This provides some evidence of stability over time.

In PISA, the tests are scaled to mean 500 and standard deviation 100 separately for each cycle, so it is not possible to directly examine changes over time. Nevertheless, the lack of a significant shift in Canada's overall international ranking indicates either that all countries have changed by similar amounts to Canada or that there has been little or no change over time. Given the stability of most national education systems, it is difficult to argue that everyone is changing by the same amount. It seems more plausible to conclude that there has been little change in either direction since PISA began.

For purposes of this paper, change in the relative position of jurisdictions is of greater interest than the overall trend over time. It is clear from these results that differences among provinces and between anglophone and francophone populations within provinces are highly persistent. While it is possible that everyone is changing (hopefully improving) at the same rate, this would be remarkable in itself. More likely, the whole system is highly stable, with little change in either absolute or relative levels of performance.

The important question for this paper is: "Why are the differences between jurisdictions so persistent?" While it might be argued that improving achievement is not explicitly stated as a goal of large-scale assessments, it would be difficult to argue that measurement serves any useful purpose if improvement is not at least an implicit goal. In particular, it would be surprising if low-performing jurisdictions did not take the results seriously enough to use as evidence that improvement is needed.

Without getting into the political debate over what these jurisdictions "should" be doing to improve achievement, it seems obvious that differences of the magnitude and persistence seen here ought to be a matter of concern. In a country with many institutions and structures explicitly devoted to equity, and particularly to minimizing disparities

across jurisdictions, the persistence of inequality in educational achievement needs to be addressed. In addition, improving the performance of the lowest-performing jurisdictions can only have salutary effects for Canadians as a whole. For example, if the performance of all jurisdictions were comparable to that for Alberta and Quebec, Canada's standing in international assessments would be even closer to the top than is now the case.

4 POPULATION (JURISDICTION/LANGUAGE) PROFILES

Overview

The most obvious starting point for investigating the sources of population differences is to identify a set of variables that show both differences across populations and positive or negative relationships to achievement. The variables of most interest are those that are amenable to change through changes in educational policies or practices, particularly at the jurisdictional level. Variables such as student demographics and socioeconomic status, which are known to influence achievement but which are not within the realm of educational policy, must also be considered, but mainly as control variables or in models that can show differential effects for different groups of specific policies or practices.

To use a simple example, we know that mother's education influences children's achievement. However, it is obvious that the school system cannot change the level of education of mothers. What might be possible, however, is for the system to help offset part of the disadvantage of low levels of mother's education, through policies and practices that focus on this particular group of children. To extend the example, we also know that the number of books in the home is positively associated with mother's education and also with achievement even after controlling for mother's education. It would not be far-fetched to consider the development of policies designed to put more books in the hands of children, particularly those with few books at home.

One of the main difficulties in modelling the variables of interest is that the focus on population differences points to populations as the main unit of analysis. Unfortunately, both theoretical and technical difficulties are encountered in conducting analysis at this high level. Theoretically, the WHW proximity model suggests that variables such as state- and district-level policies are among the lesser influences on achievement, as their impact can be felt only second-hand, through their implementation at the school or teacher level. The "fidelity of implementation" of any given high-level policy may itself be highly variable, and even the best-intentioned policies are often not implemented in the manner intended. While this is not the place to examine how jurisdictional policies are implemented, this issue imposes a significant limitation on the ability to determine how these policies influence achievement.

The technical problem relates to the unit or level of analysis. Broad jurisdictional structures and policies have as their "natural" unit the jurisdiction itself. An example would be the core curriculum or basic resource allocations. Other variables, such as teaching and learning strategies, for which the natural unit might be the school, teacher, or student, can be aggregated to the jurisdictional level for analysis. Unfortunately, using the jurisdiction as the unit of analysis yields a very small number of units from which to work and, under normal sampling assumptions, large sampling errors. Added to this, differences among the populations, though large enough to be of concern, tend to be small relative to differences among schools and students. Most of the variability in achievement is thus lost

if the analysis is conducted at the population level only. Multi-level analysis, as used in this report, only partly addresses this problem.

Initial Variable Selection

The starting point for variable selection was the *PCAP-13 2007 Contextual Report on Student Achievement in Reading* (CMEC, 2009). An initial set of variables was selected on the basis that they show statistically significant differences across populations and statistically significant relationships with achievement in either the bivariate or multivariate models presented in the Contextual Report. More specifically, for derived variables, where the scale was continuous (mean 50, SD 10), a variable was selected only if the difference between the lowest and highest jurisdiction/language group was one-half standard deviation or more.

Table 4.1 gives the set of variables initially selected, the difference between the lowest and highest population on an indicator based on the scale for the variable, and the difference in reading score between highest and lowest categories on the scale. For derived variables, such as mean factor scores, the reading score difference is between those at or below the first and at or above the fifth quintiles on the scale.

TABLE 4.1 **Initial Variable Selection: PCAP-13 2007**

Variable	Descriptive indicator as used in models	Population Difference		Difference in mean reading score from lowest to highest category of original variable	PCAP-13 2007 Contextual Report Chart Reference
		Lowest	Highest		
Student demographics					
Gender	% male students	42	54	-23	3.2, 3.4
Grade	% of students in Grade 8	56	88	71	3.6, 3.8
Language match	% of students whose home language is the same as school language	18	97	32	None
Books in the home	% with more than 200	12	34	94	3.16, 3.20
Immigrant status	% of students born outside of Canada	1	16	-13	3.22, 3.24
Teacher characteristics					
Language arts specialization	% of teachers specializing in language arts during teacher education	70	87	10	3.32, 3.34
School characteristics					
School governance	% private schools	0	33	33	3.38, 3.39
School enrolment	% more than 500	10	77	24	3.35, 3.37
Aboriginal enrolment	% of schools with more than 25% Aboriginal students	0	38	-51	3.42, 3.43
Community size	% of schools in cities more than 100,000	0	61	11	3.44, 3.45
Student reading strategies					
Reading by decoding	Mean factor score	45	53	-62	5.3, 5.7
Reading outside of class	Mean factor score	47	53	88	5.9, 5.13
Instructional climate					
Class size	% of teachers with class size 30 or more	0	60	52	6.4, 6.6
Grades in class	% of teachers with 2 or more grades in their LA classes	14	48	21	6.7, 6.8
Adjust teaching strategies to accommodate special-needs students	% of teachers who adjust teaching strategies more than a little or a lot	24	36	-36	6.19, 6.20
Time allocation and use					
Minutes per week on language arts	% of schools spending 300 or more	32	100	-16	7.1, 7.2
School average absence rate	% of schools with average absence rate more than 10%	0	33	-19	7.5, 7.6
Student absence rate	% of students reporting 11 or more days absent in school year	13	27	-29	7.7, 7.9
Teacher expected minutes per week language arts homework	% of teachers reporting one hour or more homework expected	33	68	36	7.14, 7.15

Variable	Descriptive indicator as used in models	Population Difference		Difference in mean reading score from lowest to highest category of original variable	PCAP-13 2007 Contextual Report Chart Reference
		Lowest	Highest		
Homework counts toward marks or grades	% of teachers reporting that homework "often" counts toward marks or grades	11	75	-16	7.17, 7.18
Student homework in all subjects	% of students with 2 hours or more homework per week	25	50	59	7.19, 7.22
Tutoring	Student is being tutored	18	47	-34	5.14
Teaching strategies in reading					
Creative reading materials	Mean factor score for teacher use of creative reading materials	44	56	23	8.9, 8.10
Assignment of reading outside of class	% of teachers who "often" assign reading outside of class	18	60	24	8.11, 8.12
Re-teach basic reading skills	% of teachers who "often" re teach basic reading skills	21	50	-19	8.13, 8.14
Use of media	Mean factor score for teacher use of media in reading	46	54	-62	8.15, 8.20
Assessment					
Influence of external exams	% of principals who "strongly agree" that external exams influence school program	12	59	13	Not in Contextual Report
Student reported assessment by short test items	Mean factor score	45	54	-38	9.1, 9.5
Teacher use of short test items	Mean factor score	46	54	-31	9.6, 9.8
Teacher use of long test items	Mean factor score	48	53	17	9.7, 9.8
Teacher use of student assignments/projects for assigning grades	% of teachers using 1–2 times a year or more	47	97	-28	9.11, 9.12
Teacher use of non-academic criteria in assigning grades	% of teachers using 4–5 such criteria	7	50	-51	9.13, 9.14
Knowing what a rubric is	% of students who know what a rubric is	27	76	21	9.16, 9.19
Use of rubrics for marking	% of students whose teachers "often" use a rubric for marking	16	61	33	9.17, 9.19
Teacher assessment skills	% of teachers who are "very skilled" at assessment	32	83	15	9.20, 9.21

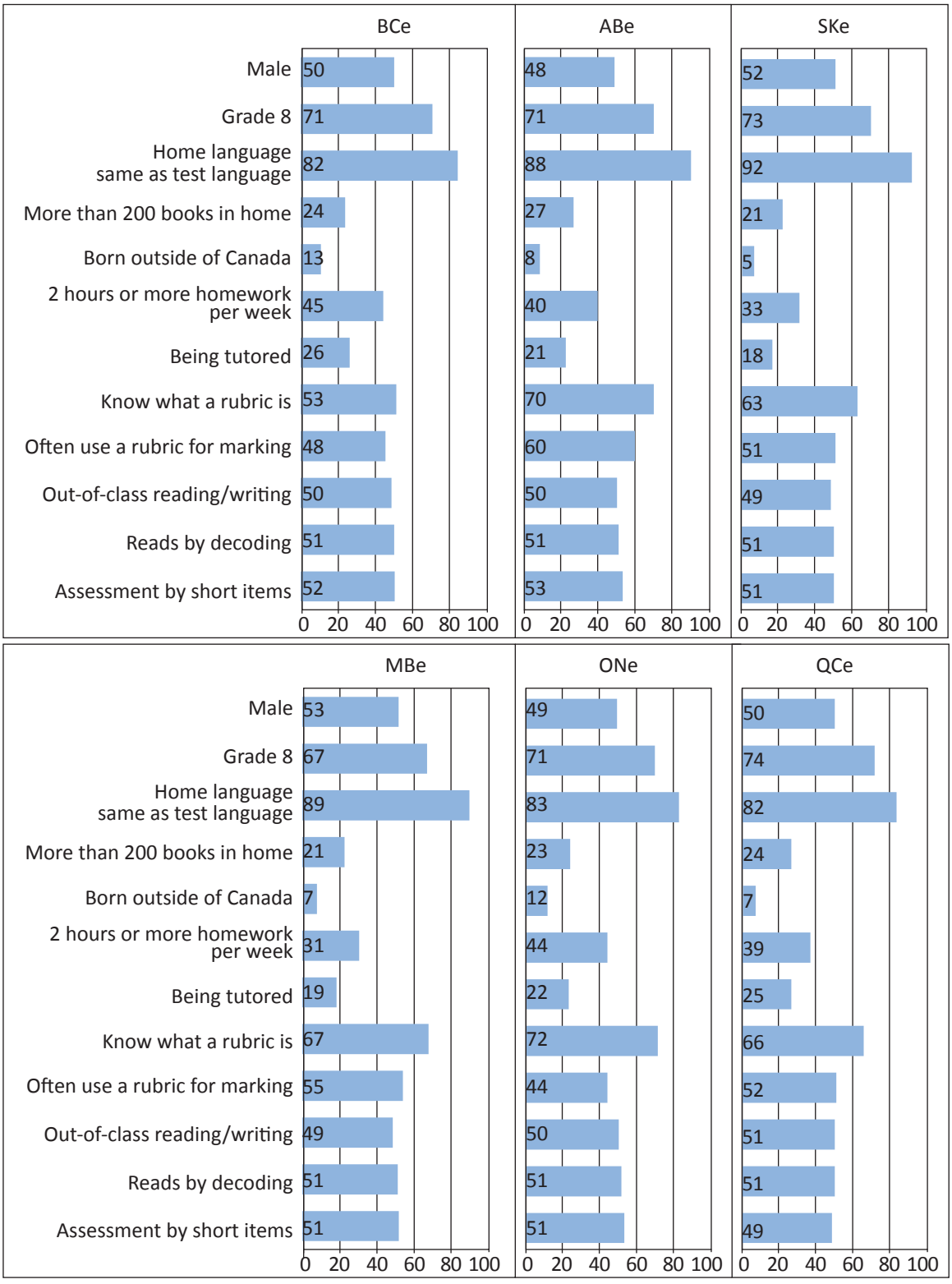
Unfortunately, there is no straightforward way to draw inferences about population differences directly from this table. The usual approaches to this problem are either to examine population differences one variable at a time, or to model the combined effects of many variables one population at a time. Neither of these is appropriate when the goal is to determine if there are consistent patterns on variables related to achievement that can be used to explain population differences.

In this report, we take two alternative approaches. The first is to develop "population profiles" based on the position of each of the populations of interest on each of the variables given in Table 4.1. The second involves modelling achievement using the variables indicated, along with the populations, coded as binary variables (dummy coded) in the models. The first approach is detailed in this chapter and the second in the next chapter.

Student Profiles

The student variables given in Table 4.1 were used to develop graphical profiles of students in each population. Chart 4.1 shows these for English populations and Chart 4.2 for French populations. The intention here is to give an "at-a-glance" comparative picture of the student populations across the jurisdictions and language groups.

CHART 4.1 **Student Profiles: English Populations**[6]



[6] Minor differences may be found between the percentages given in this chapter and those in the PCAP-13 2007 Contextual Report because of differences in the treatment of missing data. None of these differences affect the overall pattern of results.

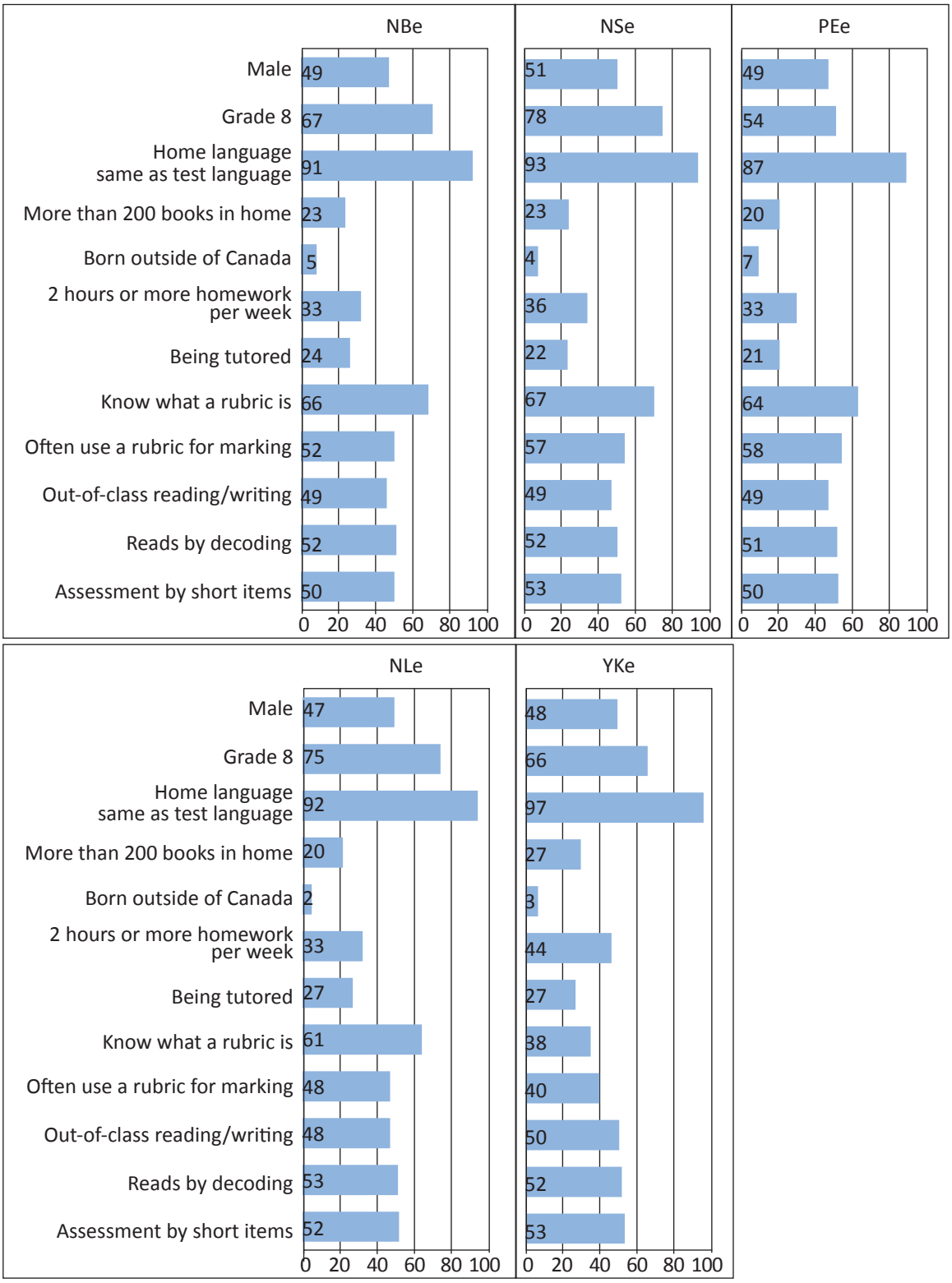
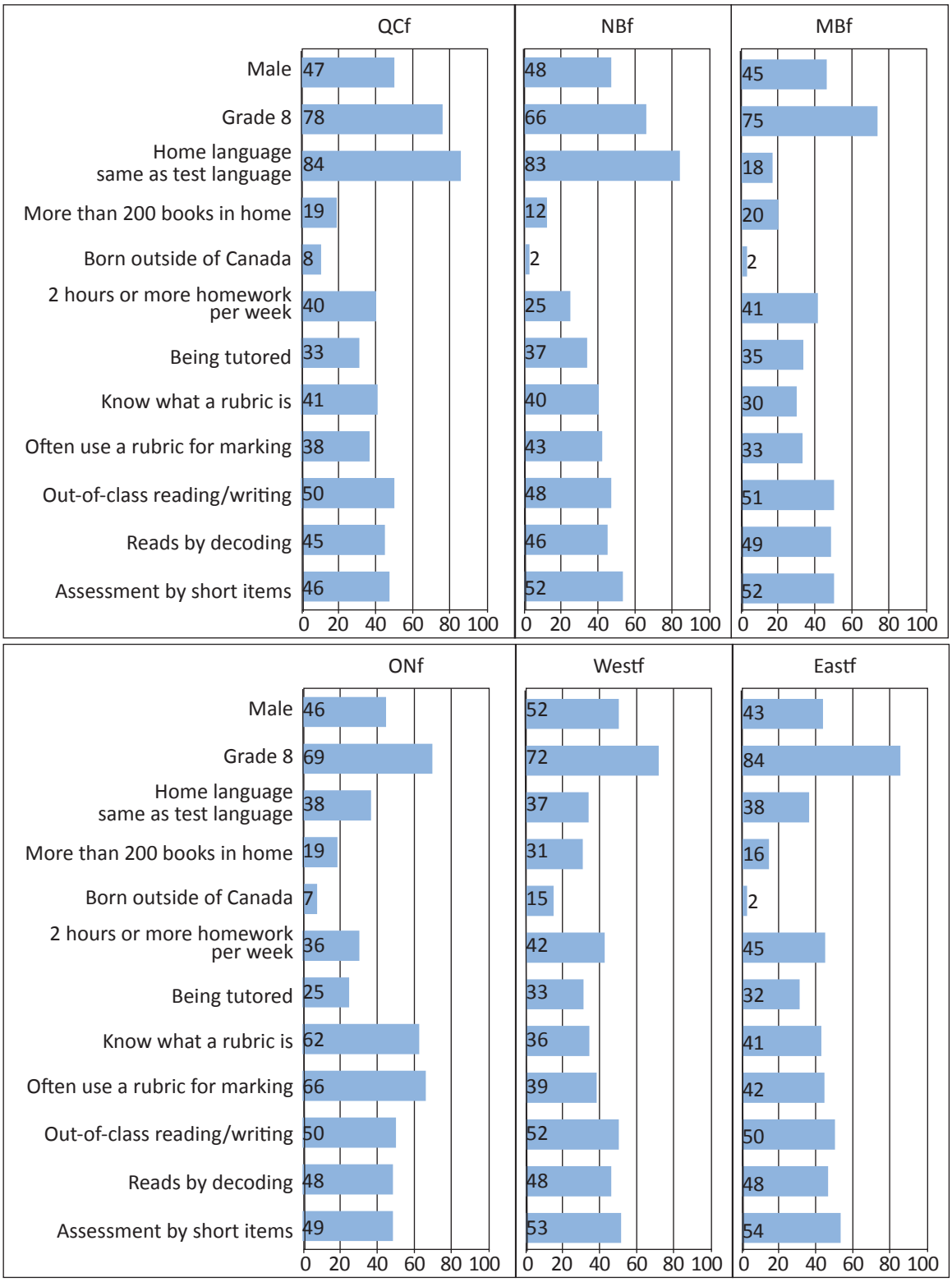

NBe
NSe
PEe
NLe
YKe
Male
Grade 8
Home language same as test language
More than 200 books in home
Born outside of Canada
2 hours or more homework per week
Being tutored
Know what a rubric is
Often use a rubric for marking
Out-of-class reading/writing
Reads by decoding
Assessment by short items
49
67
91
23
5
33
24
66
52
49
52
50
51
78
93
23
4
36
22
67
57
49
52
53
49
54
87
20
7
33
21
64
58
49
51
50
47
75
92
20
2
33
27
61
48
48
53
52
48
66
97
27
3
44
27
38
40
50
52
53
0 20 40 60 80 100

CHART 4.2 **Student Profiles: French Populations**

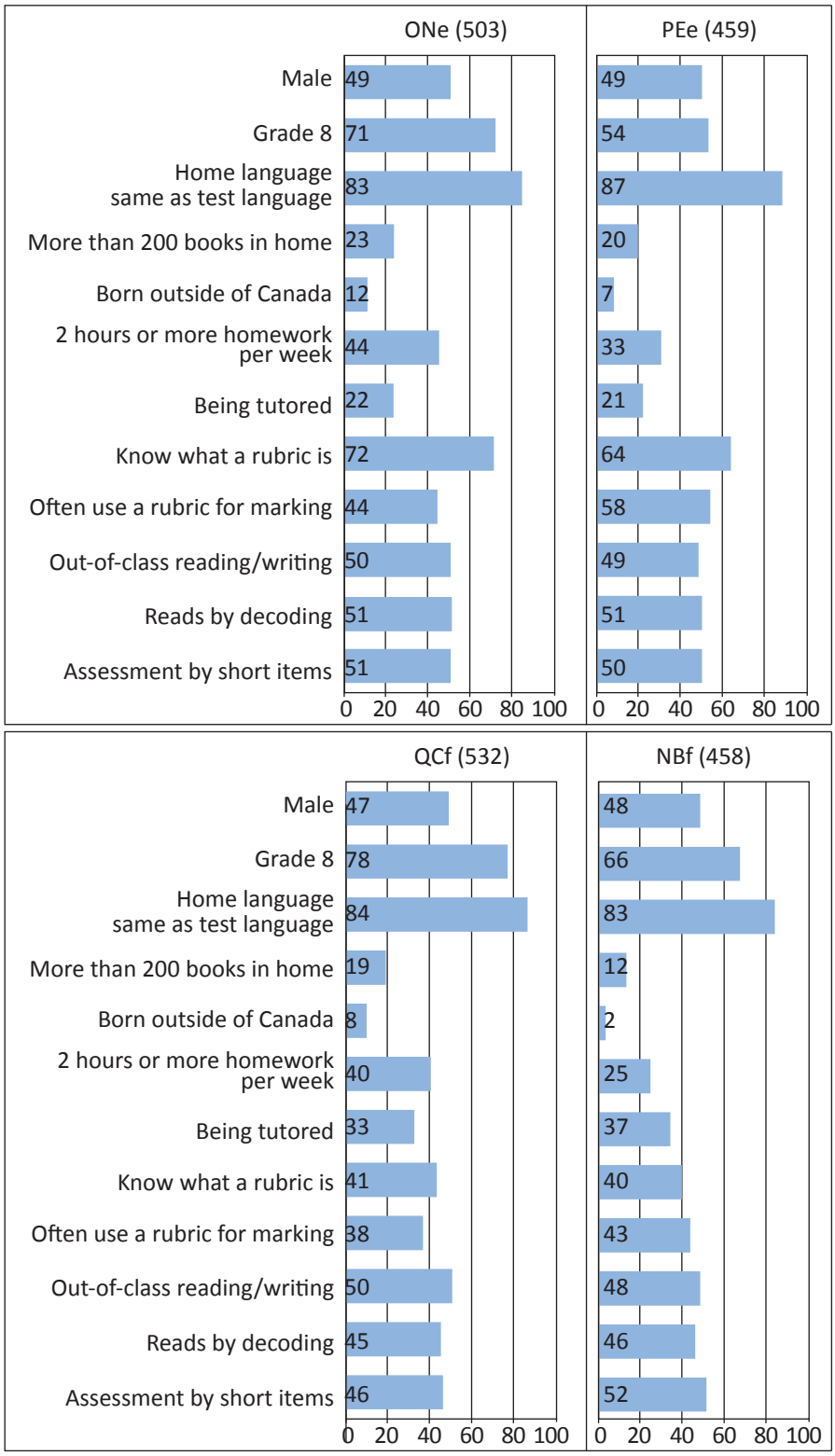


It is immediately apparent from these graphs that student characteristics are more alike than different across populations. Although all of these variables are significantly associated with achievement and show statistically significant differences across populations, the profile graphs suggest that the latter are not large enough to have much of an impact on overall achievement levels. The implication is that, while student-level variations in these characteristics are large enough to influence the achievement levels of individual students, the population-level differences are likely not large enough to have much of an impact on population-level achievement.

Although the broad picture of population differences is apparent from these graphs, the large number of populations and variables makes it difficult to discern much detail in the differences. An alternative way of looking at the results is to look more closely at the populations that are at the extremes of the achievement scale. Chart 4.3 shows these results for the extreme populations for both languages.[7]

[7] New Brunswick French (reading score 458) is used rather than Manitoba French (reading score 436) because the Manitoba French population includes French immersion students and is thus not directly comparable to other French populations.

CHART 4.3 **Student Profiles for Populations with Highest and Lowest Reading Scores**



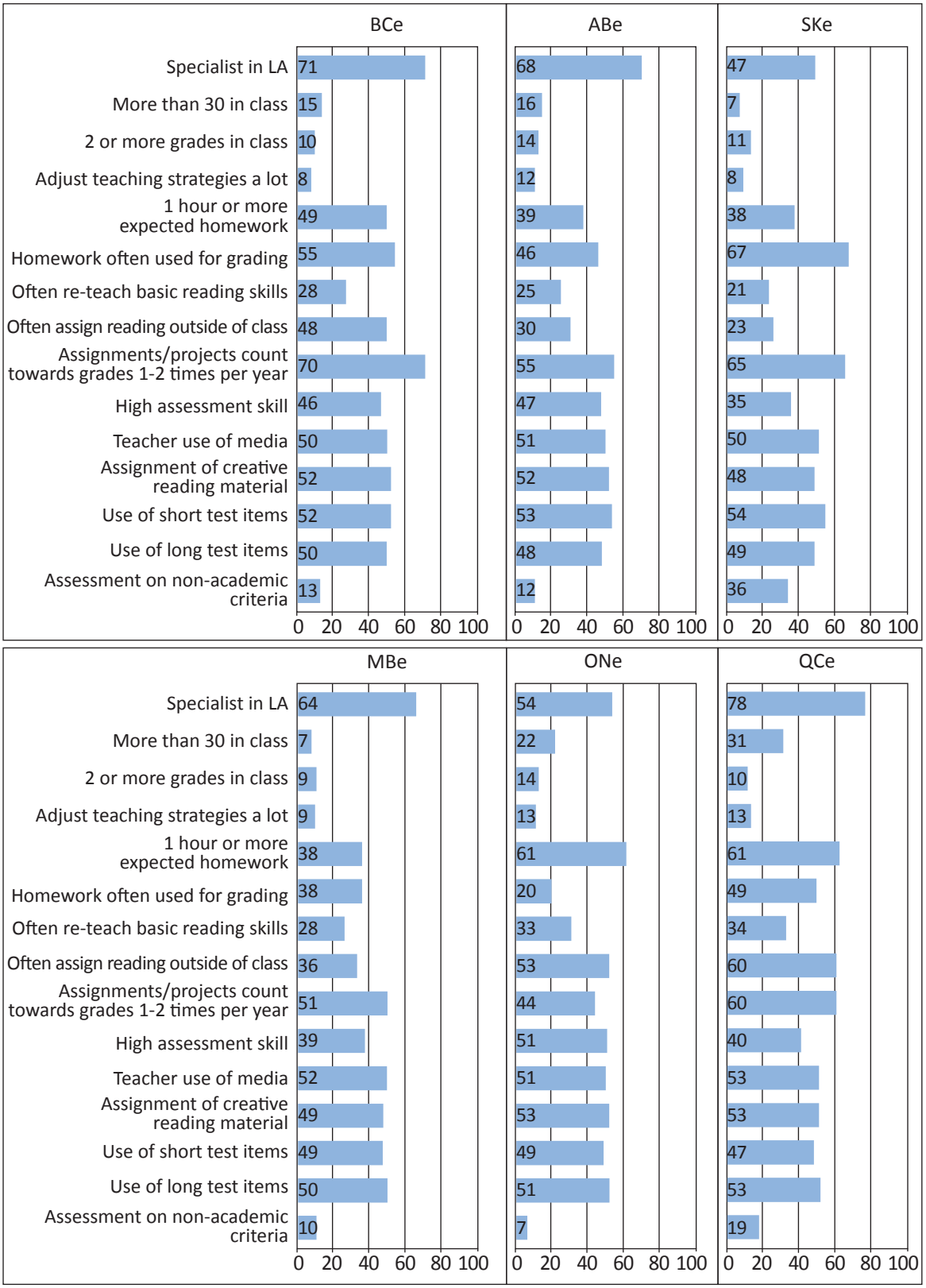
These graphs show some differences worth noting for their achievement effects. The main ones are:

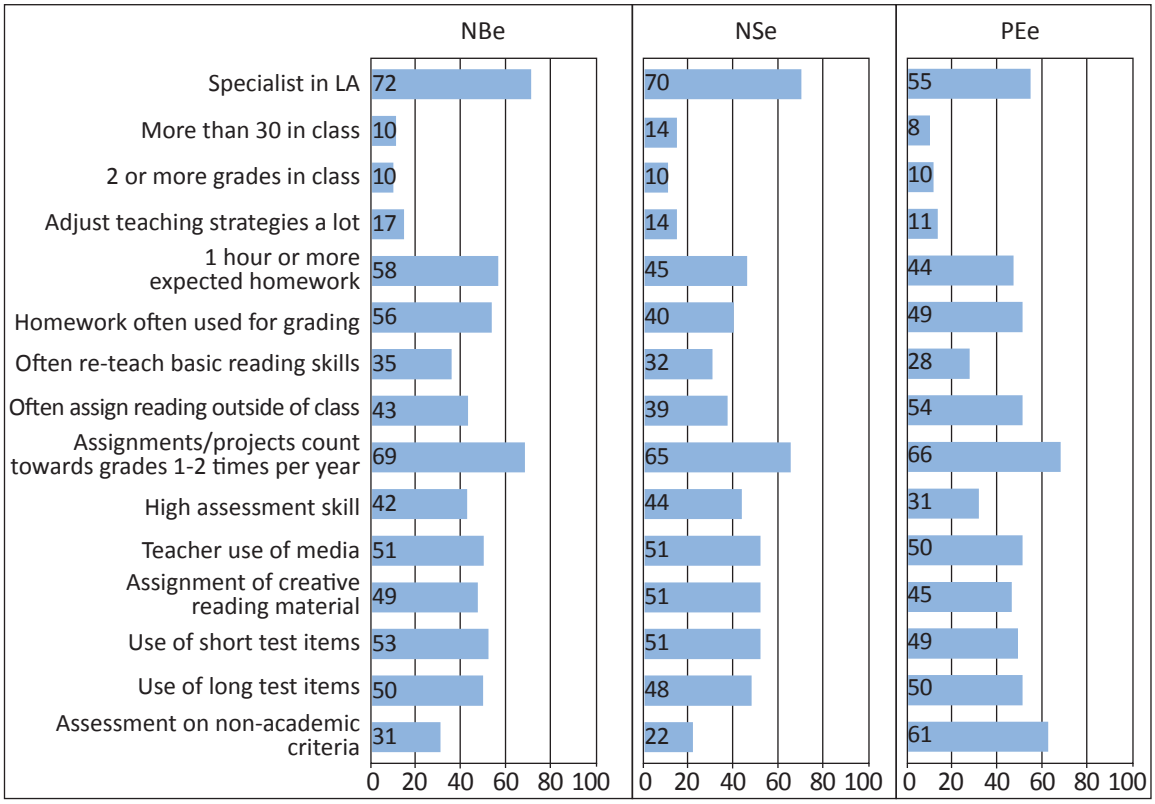
- Both of the highest-performing populations (Ontario English and Quebec French) have more students in Grade 8 than the lowest-performing provinces (Prince Edward Island English and New Brunswick French). Both of the latter actually have more students in Grade 9 than any other population. On average, those in higher grades have higher reading scores. This indicates that the average score for the lower-performing populations would have actually been lower had that population had the same proportion of Grade 8 students as the two highest-performing populations. However, a direct comparison of populations for Grade 8 students only showed that the averages change little under that restriction.
- Similarly, both of the highest-performing populations have more students who are born outside of Canada compared to the lowest-performing populations. Again, since being born outside of Canada is negatively associated with reading performance, this should have the effect of depressing the scores for the higher-performing populations. Obviously, this negative effect is being offset by other positive effects.
- An example of such a positive effect is homework. Homework is positively associated with achievement, and students in the highest-performing populations tend to do more homework than those in the lowest-performing populations.
- For the francophone populations, there is also a difference in the number of books in the home. This difference is in the expected direction, with the Quebec francophone population having more books than the New Brunswick francophone population.
- For the anglophone populations, more students in Ontario English know what a rubric is than is the case for Prince Edward Island English. This is in the expected direction, since knowing what a rubric is, is positively associated with achievement. Although more students in Prince Edward Island English reported that they use a rubric for scoring, this is actually a proportion of those knowing, and hence is actually lower in absolute terms than the percentage in other populations.

Teacher Profiles

Chart 4.4 and 4.5 give population profiles based on the teacher variables in Table 4.1. Again, the general picture is one of greater similarities than differences among the populations. However, a few larger differences are evident, which again are best illustrated by looking at the highest- and lowest-performing populations. These are shown in Chart 4.6.

CHART 4.4 **Teacher Profiles: English Populations**




NBe
NSe
PEe
Specialist in LA
More than 30 in class
2 or more grades in class
Adjust teaching strategies a lot
1 hour or more expected homework
Homework often used for grading
Often re-teach basic reading skills
Often assign reading outside of class
Assignments/projects count towards grades 1-2 times per year
High assessment skill
Teacher use of media
Assignment of creative reading material
Use of short test items
Use of long test items
Assessment on non-academic criteria
72 10 10 17 58 56 35 43 69 42 51 49 53 50 31
70 14 10 14 45 40 32 39 65 44 51 51 51 48 22
55 8 10 11 44 49 28 54 66 31 50 45 49 50 61
0 20 40 60 80 100

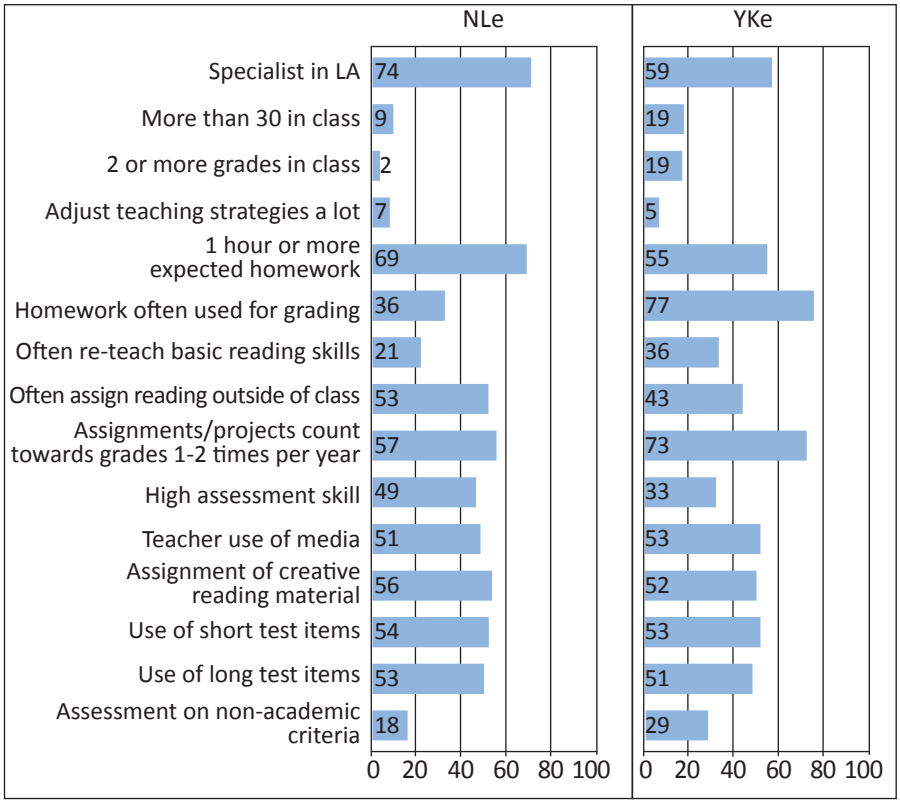
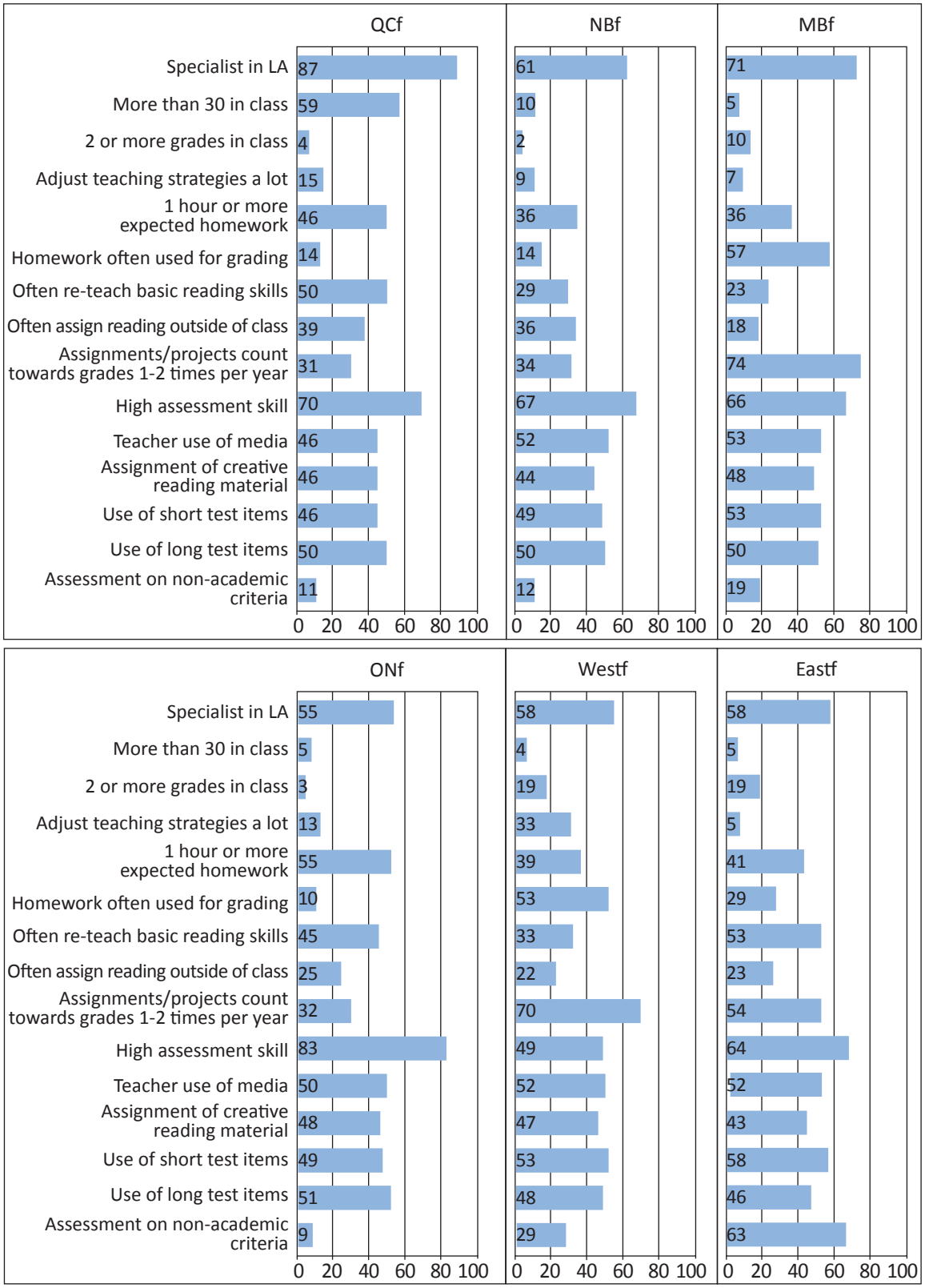

NLe
YKe
Specialist in LA
More than 30 in class
2 or more grades in class
Adjust teaching strategies a lot
1 hour or more expected homework
Homework often used for grading
Often re-teach basic reading skills
Often assign reading outside of class
Assignments/projects count towards grades 1-2 times per year
High assessment skill
Teacher use of media
Assignment of creative reading material
Use of short test items
Use of long test items
Assessment on non-academic criteria
74 9 2 7 69 36 21 53 57 49 51 56 54 53 18
59 19 19 5 55 77 36 43 73 33 53 52 53 51 29
0 20 40 60 80 100

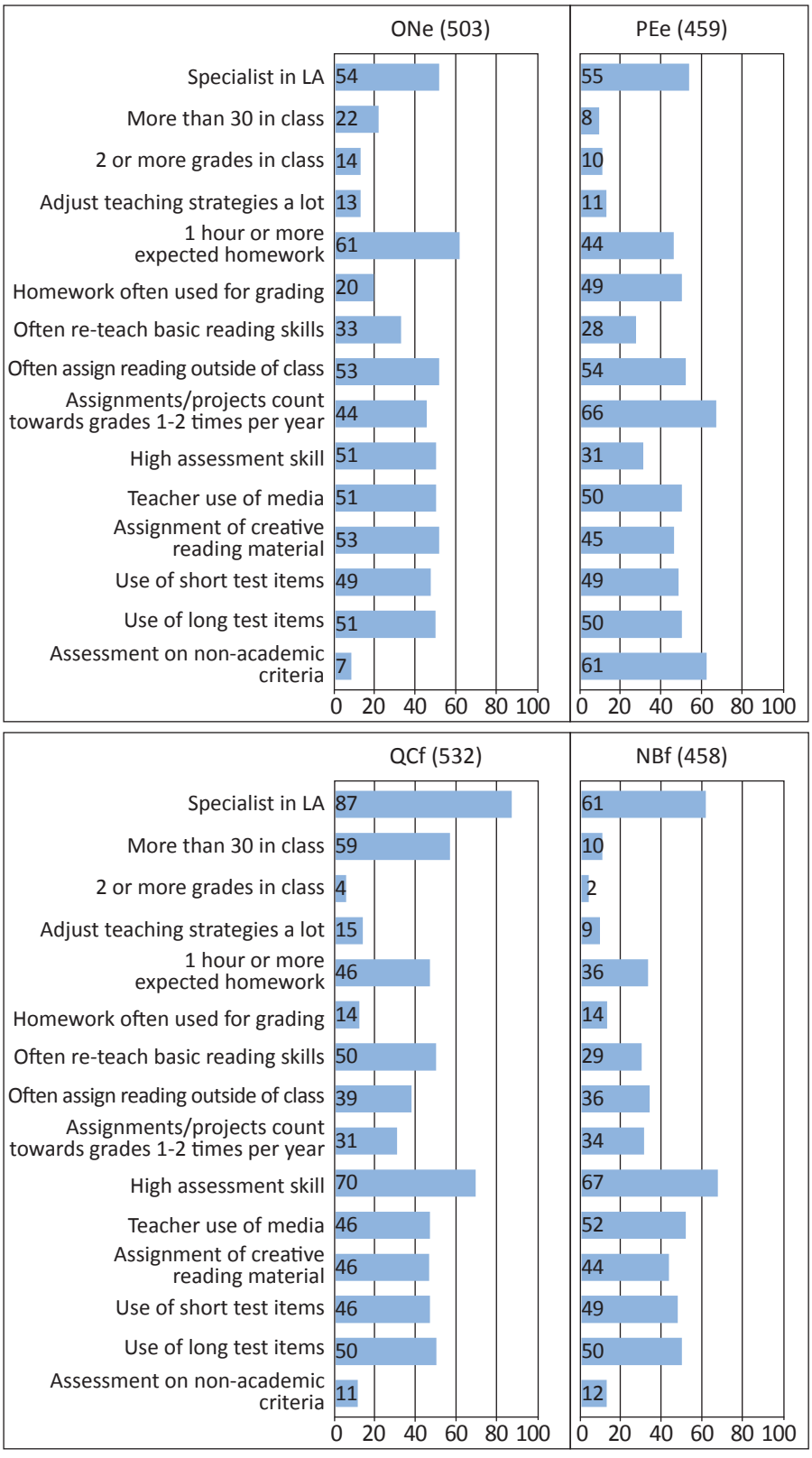
CHART 4.5 **Teacher Profiles: French Populations**



The main differences highlighted by Chart 4.6 are:

- Both of the high-performing populations have higher percentages of classes with more than 30 students than the low-performing populations. While this seems counterintuitive, this is consistent with the overall finding that students in larger classes have higher performance levels. This is true even after controlling for other variables related to class size (CMEC, 2009).
- Teachers in the two high-performing populations expect more homework from their students than is the case for the low-performing populations.
- Teachers in Prince Edward Island English use homework for grading more than those in Ontario English. However, there is no significant difference between the two francophone populations on this variable.
- Quebec francophone teachers more often re-teach basic reading skills than their New Brunswick francophone counterparts. However, this variable is negatively related to achievement, and thus cannot help account for the higher achievement of Quebec francophone students.
- Higher assessment skill is positively related to achievement. Ontario English teachers reported having higher levels of assessment skill than Prince Edward Island English. However, there is no significant difference between the two francophone populations on this variable.
- Using non-academic criteria for grading is negatively related to achievement. This occurs much more often for Prince Edward Island English teachers than for either their Ontario English counterparts or those in either of the francophone populations.

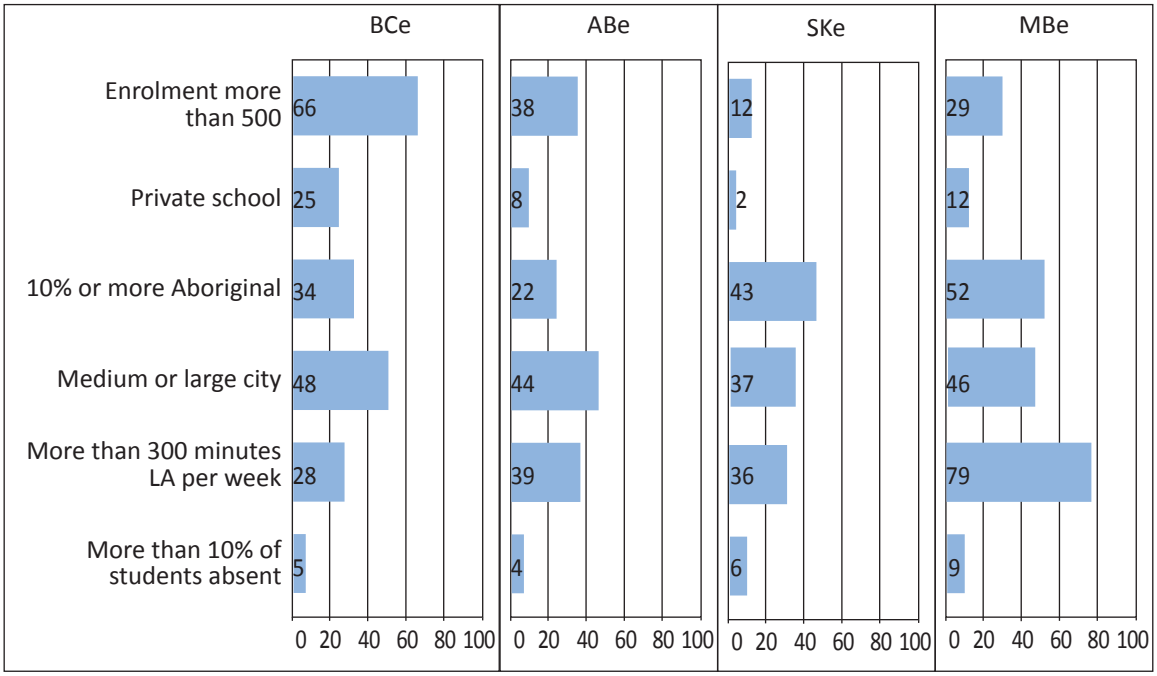
CHART 4.6 **Teacher Profiles for Populations with Highest and Lowest Reading Scores**



School Profiles

Charts 4.7 and 4.8 give population profiles based on the school variables in Table 4.1. Again, there are some differences that can be examined by looking at the populations at both extremes of the achievement scale. These are shown in Chart 4.9. However, one difference should be noted that is not evident from the extremes. Chart 4.7 shows that all of the schools in the Yukon population have more than 10% students of Aboriginal identity. The proportion of schools in this category is also high for Saskatchewan English and Manitoba English and French.

CHART 4.7 **School Profiles: English Populations**



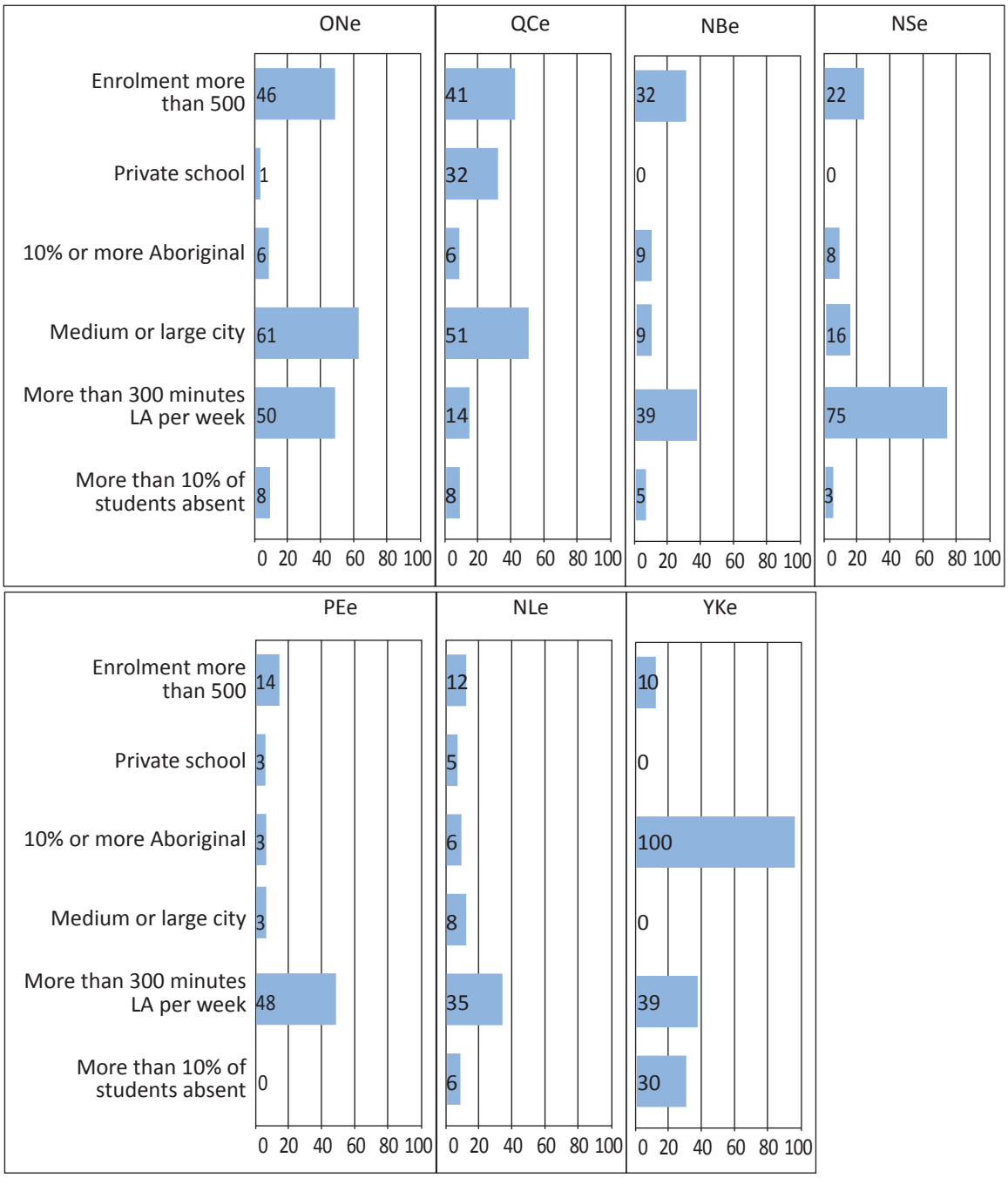
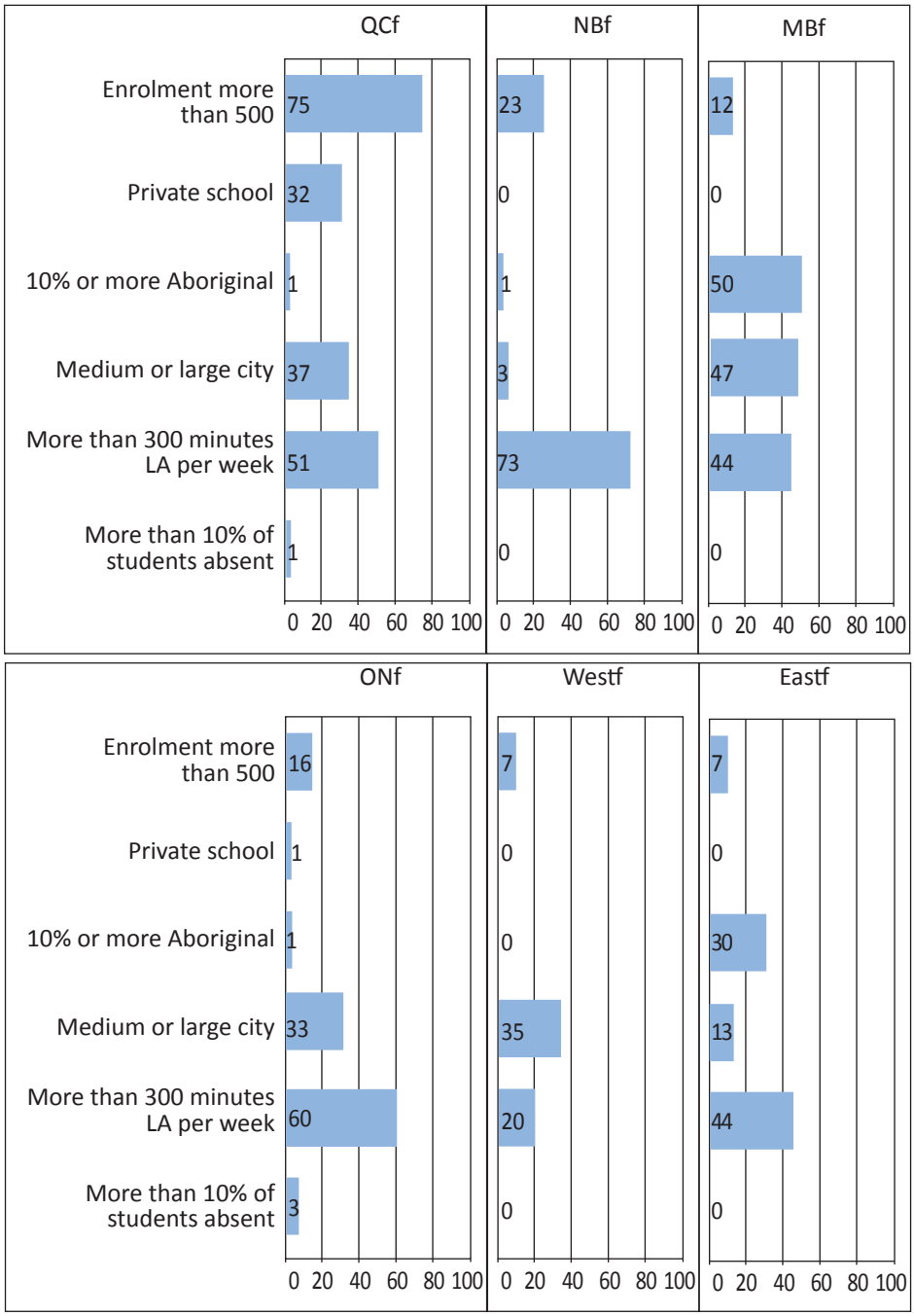

ONe
QCe
NBe
NSe
PEe
NLe
YKe
Enrolment more than 500
Private school
10% or more Aboriginal
Medium or large city
More than 300 minutes LA per week
More than 10% of students absent
46
1
6
61
50
8
41
32
6
51
14
8
32
0
9
9
39
5
22
0
8
16
75
3
14
3
3
3
48
0
12
5
6
8
35
6
10
0
100
0
39
30
0 20 40 60 80 100

CHART 4.8 **School Profiles: French Populations**

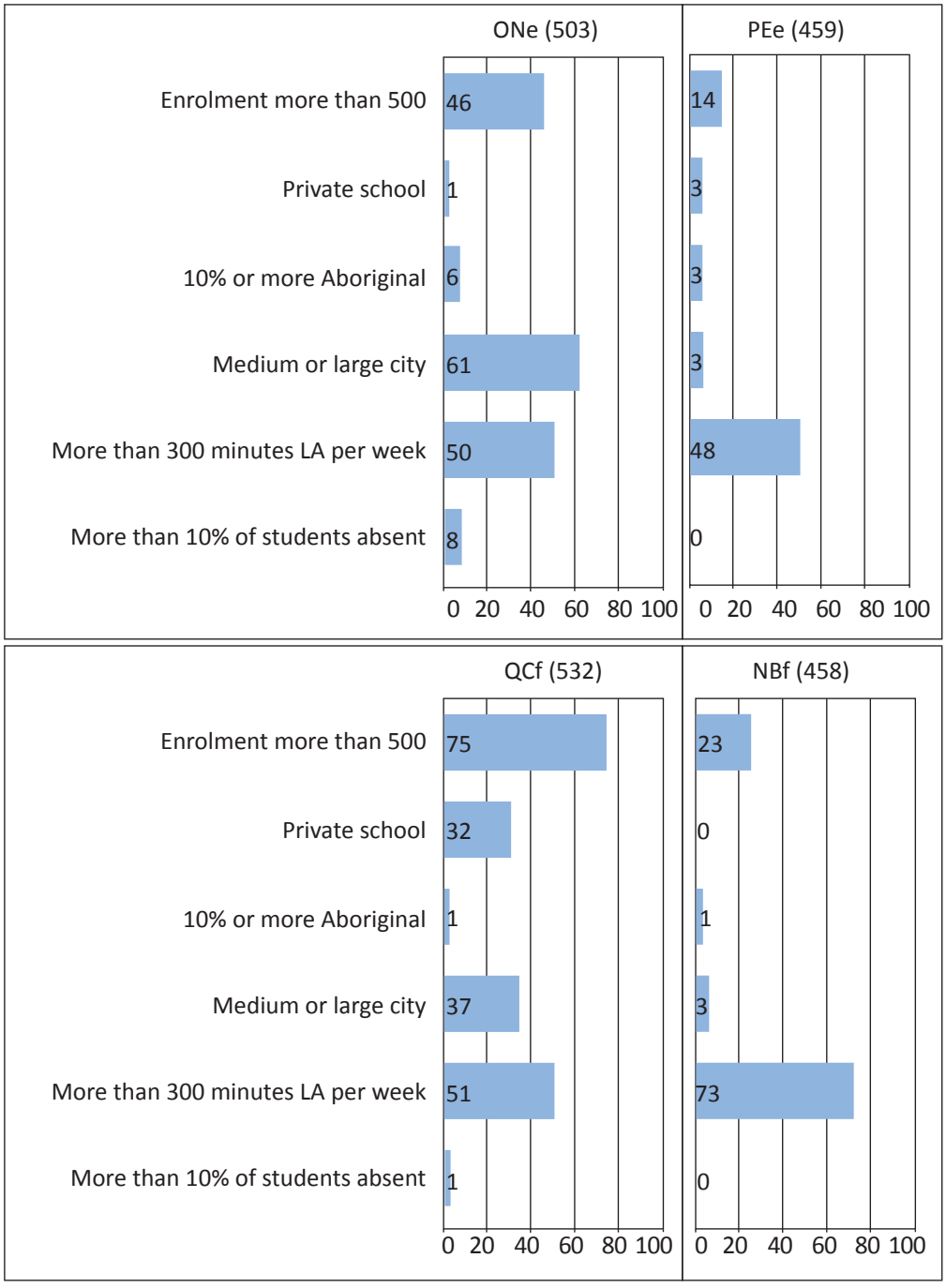


As for Chart 4.9, the main differences are:

- More schools in Ontario English and Quebec French, the high-performing populations, are larger and are in larger communities than in the low-performing populations.
- In Quebec French, there are more private schools than in other populations. Students in private schools, especially in Quebec, have higher reading achievement than those in public schools, even after controlling for other variables (CMEC, 2009). The private school effect is thus a plausible factor in accounting for the exceptionally high performance of Quebec francophone students.

- New Brunswick francophone schools (the lowest-performing francophone populations) are more likely to spend more than 300 minutes per week on language arts than is the case for other populations. This variable is negatively related to achievement, which suggests that schools in which students are not performing well in language spend more of their time on that subject without this having the desired effect on achievement. The problem here, of course, is that there is no way to tell whether spending more time has a positive effect on those who are given the added time. What is clear is that the added time is not having a decisive effect in overcoming lower average achievement.

CHART 4.9 **School Profiles for Populations with Highest and Lowest Reading Scores**



5 MODELLING ACHIEVEMENT DIFFERENCES

Multi-Level (Hierarchical) Modelling: Population, Intermediate and Full Models

In this chapter, the PCAP-13 2007 reading scores at the individual level are treated as dependent variables, and the variables described in the previous section are used as predictors of these scores. The usual method of analysis in this situation is multiple regression analysis. Multi-level or hierarchical modelling is a variation on multiple regression analysis, used when the sampling design is a hierarchical one. In this case the hierarchical design arises because students are "nested" within schools, and schools are nested within populations. Thus, the models may be two-level (students within schools) or three-level (students within schools within populations). Two-level models are used in this report because the number of populations is too small to permit three-level analysis.

The models used in this chapter were developed in stages as follows:

- The overall variation in student achievement was partitioned into proportions attributable to differences between students and differences between schools. This helps determine if differences between schools are sufficient to justify using schools as a separate level of analysis.
- A variable was created for each PCAP population, and each student in the data file was coded as 0 or 1 to identify his or her membership in a population.
- The initial model, referred to as a "population model" consisted of all of the populations, based on the 0/1 coding, entered into the model as a group of predictors, with reading score as the dependent variable.
- The regression coefficients for this model correspond to the difference in reading achievement between each population and a reference population, in this case Ontario English. A reference population is required to avoid a "linear dependency," which prevents the model from being computed. To obtain a coefficient for Ontario English, a separate model was run with Quebec French as the reference population.
- Other variables were then entered into the model in a sequence determined by either theoretical or empirical criteria. The coefficients of these models may be expressed as the change in reading score associated with a one-unit change in the predictor, as other variables are controlled.
- Once all of the predictor variables are entered, the model is referred to as a "full" model. This model gives the "unique" effect of a particular predictor, with all other predictors controlled. The full model is the one of most interest in the analysis. Comparing the coefficient for a population in the full model with that for the same predictor in the population model allows us to examine the effect of the full predictor set on achievement in that population. Differences in these effects may thus be interpreted as differences in how the predictor set affects reading achievement in each population separately.

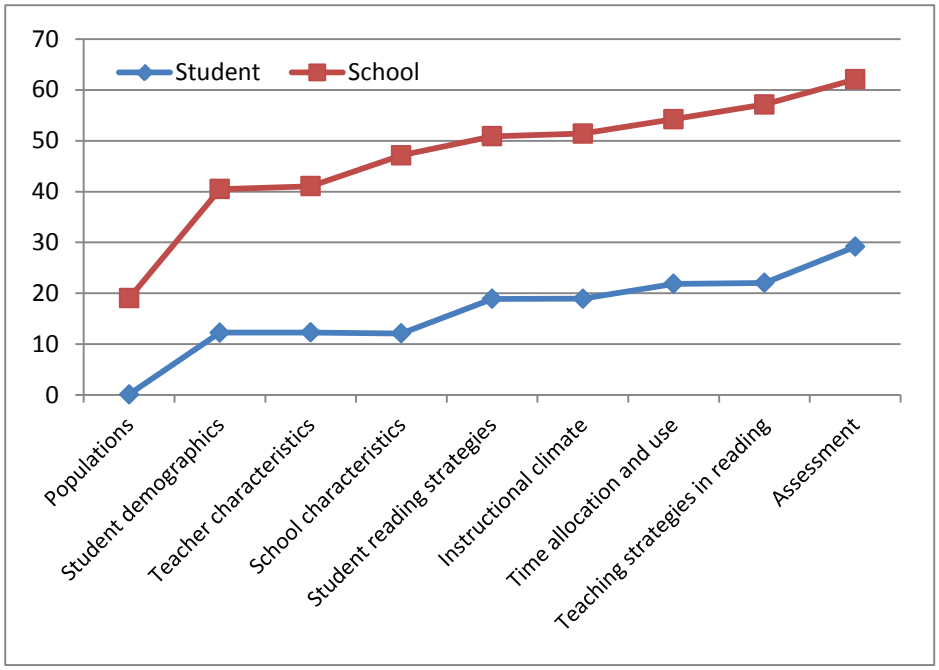
Proportions of Variance

One of the main features of hierarchical modelling is that it allows us to determine how much of the total variation in the outcome is attributable to differences between students and between schools. To determine the proportions of school and student variance, an initial or "null" model is estimated, in which no independent variables are entered. This model produces the total variance in the dependent variable as well as the between-student and between-school proportions of this total. In this case, the null model shows that differences between students account for 85% of the variance in reading scores, leaving 15% to be accounted for by differences between schools.

Taking the total student and school variance as the starting point, Chart 5.1 shows how the proportions of variance at each level change as clusters of variables are added to the model. It is important to note that the variable clusters have been entered cumulatively, so the variance explained by each cluster is added to that explained by all of the previous clusters.

The first cluster is the set of population variables, as described in the text box above. The remaining clusters correspond to those shown in Table 4.1. It is important to note that the figures shown in the chart are "proportions of proportions." For example, entering the student demographic variables accounts for about 12% of the original 85% that is student variance (that is, 12% of 85%) and 40% of the school variance (40% of the original 15%). Because this may seem somewhat complex, the important point to note is that the proportions of variance accounted for by the model increase as more variables are added. After entering all of the selected variables, the model accounts for a much larger proportion of the school variance than of the student variance. That is, the model is more effective in explaining differences between schools than differences between students.

CHART 5.1 **Proportions of Variance Accounted for by Model Stages**



Looked at in more detail, Chart 5.1 shows that the largest proportional increase in explained student variance occurs when student demographics are entered into the model. Relatively large changes in explained student variance also occur for school demographics and assessment, indicating that these clusters are the ones most strongly influencing student achievement. Student characteristics also yields a large increase in the proportion of explained school variance, indicating that much of the variation between schools is accounted for by variations in the backgrounds of students in the school. Beyond this, explained school variance increases slightly and fairly steadily as other variable clusters are added.

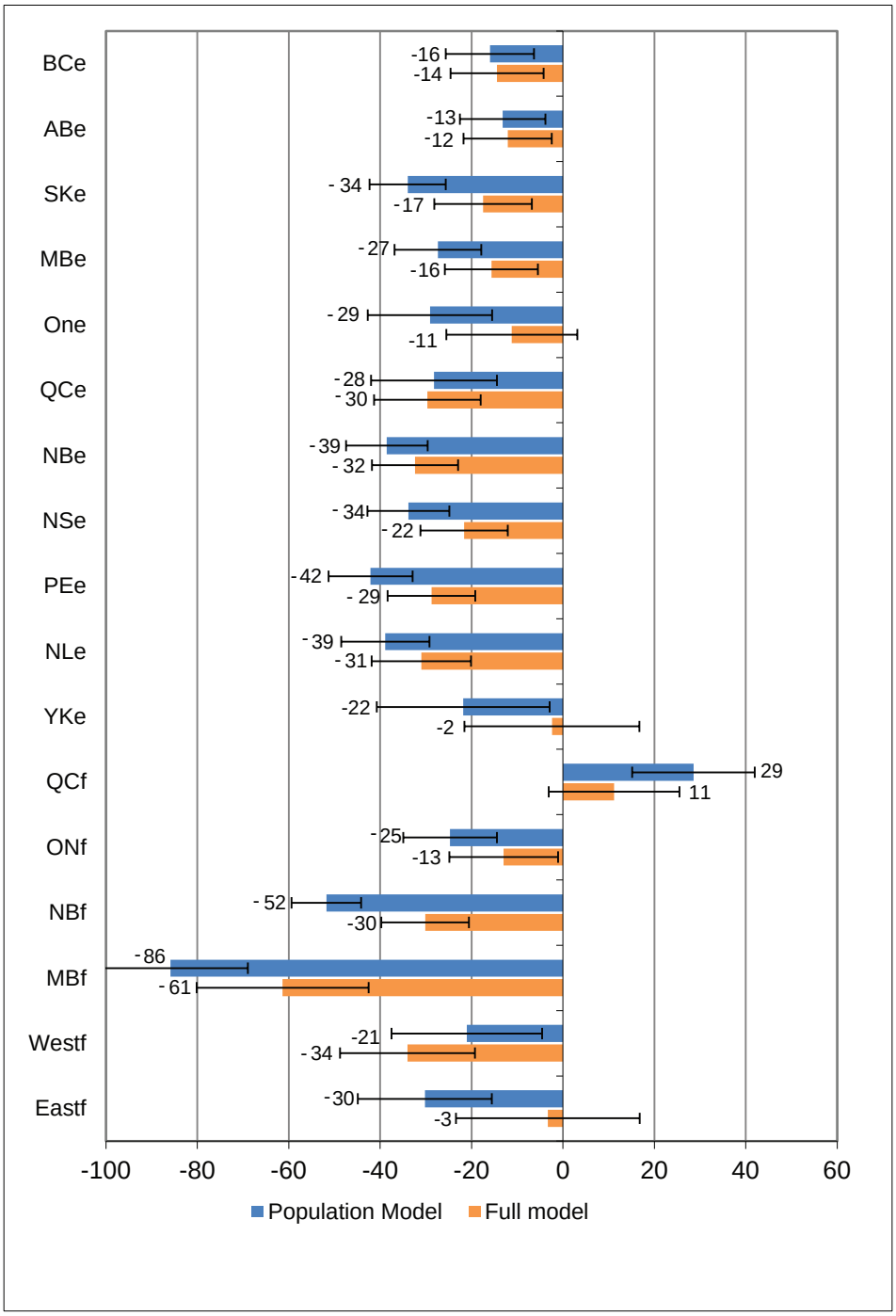
Population Effects

The PCAP-13 2007 Contextual Report (CMEC, 2009) presented models for the effects of a variety of variables on reading achievement. The concern here is not directly with these effects but rather with how these may act to influence differences in achievement levels across populations. The starting model for this report was thus one in which the populations were entered as independent variables, with each population "dummy coded" (coded as 0 or 1). In this case, one population had to be omitted because including all creates a "linear dependency," which prevents the model from converging to yield the desired estimates. This population then serves as a "reference group" against which all other populations are compared. In this case, Ontario English was treated as the reference group because this was the population with average achievement closest to the Canadian average. In order to obtain coefficients for Ontario English, a second version of all of the models was run with Quebec French as the reference group and Ontario English included. The latter model yields different population coefficients but does not affect the ability to examine the main issue, that of how the added variable clusters change the population coefficients.

Because entering each variable cluster requires a new model, the total set of models yields a large number of coefficients. For that reason, and because it is useful, first, to examine the overall extent of change, only the results for the population and full models are shown in Chart 5.2. The full set of coefficients is given in Appendix A (Tables A.1 and A.2.)[8]

[8] The display in Chart 5.2 follows the convention established in earlier PCAP reports. The lines on each bar are "error bars" representing the confidence interval associated with each coefficient. Differences between the initial and full models or between populations are considered to be statistically significant if the error bars do not overlap. The confidence intervals are derived from the standard errors produced by the model, and are adjusted for "finite populations," where the sample size approaches the population size in some populations.

CHART 5.2 **Population Coefficients for Population and Full Models**



In this chart, except for Ontario English itself, the coefficients for the population model represent the difference in the PCAP reading score between Ontario English and each of the other populations. For Ontario English, the population model coefficient is the difference between that population and Quebec French. All of the coefficients are negative, with the exception of Quebec French, reflecting the fact that Ontario English has a higher reading score than all other jurisdictions except for Quebec French.

The important feature of this graph is actually the change in coefficients from the population model to the full model. While these appear to undergo large shifts in some cases, the shift is statistically significant for only one population, New Brunswick French. With only a couple of exceptions, the full model coefficients also remain significantly negative. All of this indicates that, in general, this set of variables does not change the pattern of differences among populations in any significant way. Thus, whether or not a population is high or low on the variables in the model is generally not a significant contributor to that population's reading performance relative to that for other populations.

Closer examination of the coefficients for the intermediate models for New Brunswick French (Appendix A, Table A.1), shows some important trends. A statistically significant shift in the coefficient (from -51.78 to -37.82, $p<0.05$) for that population is found when student demographic variables are added to the model. This shows that controlling for student demographics (gender, grade, born outside of Canada, language match, and number of books in the home) reduces the difference between New Brunswick French and the reference group, Ontario English. Within this group of variables, the largest effects on achievement are for grade and language match. Controlling for the grade effect should make little difference to the comparison because New Brunswick French and Ontario English have similar distributions of students across the grades. The same is true for language match, where both of these populations have a greater than 80% match of home language to school language. The greatest source of contrast is for books in the home, where Ontario English students have, on average, many more books than New Brunswick French students. The latter variable is the better of our two measures of socioeconomic status (the other is mother's education), thus suggesting that part of the shift in coefficients for New Brunswick French is related to socioeconomic status.

Interestingly, the coefficient for New Brunswick French begins to increase again as school characteristics, teacher characteristics, and student reading strategies are added to the model. This suggests that some of the variables in this category are having a negative effect on the performance of this population. None of these variables has a particularly strong effect on achievement, and some have negative and others positive effects. It is therefore difficult to attribute the observed change to any specific variables within these clusters. The coefficients actually decrease again (become less negative) as the remaining clusters are added to the model, indicating that instructional climate, time, teaching strategies, and assessment strategies and practices have overall positive effects on New Brunswick French students.

The above discussion illustrates the complexity of interpretation required to account for the observed shifts in coefficients. In effect, this indicates that, in addition to these variables having relatively small effects overall, they interact in complex ways to produce many of the shifts that are found to occur. Much of this complexity is due to the fact that these variables are correlated with each other, as well as with achievement. Variables can thus exert independent effects, when taken alone or combined, or interacting effects when taken together.

This point can be illustrated further by reference to a few other statistically significant changes that are found in the intermediate models, which do not show up in the full model because of these complex interaction effects. Examples are:

- For Saskatchewan English, the coefficient changes significantly (-31.20 to -22.35 $p<0.05$) when school characteristics are added to the model. The most likely source of this shift is that adding this cluster controls for the relatively large number of Saskatchewan English schools with a high proportion of Aboriginal students.
- The same is true for Ontario English (-38.06 to -23.99 $p<0.05$). Since this shift is relative to Quebec French, the most likely source of the shift is that this is a consequence of controlling for the large positive private school effect for Quebec French.
- For Quebec French, adding the school characteristics changes the coefficient from 38.02 to 23.99 ($p<0.05$). The most obvious source of this shift is the large private school effect for this population. However, other positive school characteristics, including school size and class size, are also favourable to the Quebec French population.

6 ACHIEVEMENT EQUITY

The argument is commonly made that jurisdictions should strive not only for high average achievement but also for greater equity in achievement. The most recent PISA Canada report (Knighton, Brochu, & Gluszynski, 2010) indicates that Canada is one of the few PISA countries to show both high achievement and high equity. Variations in equity across provinces are also relatively small. Willms (2003) and others have argued that a desired goal of schooling is reducing the degree of inequality across socioeconomic groups. However, from an educational policy perspective, other areas of inequality, such as those between genders, urban–rural schools, or quality of school programs or teaching strategies, are also of interest.

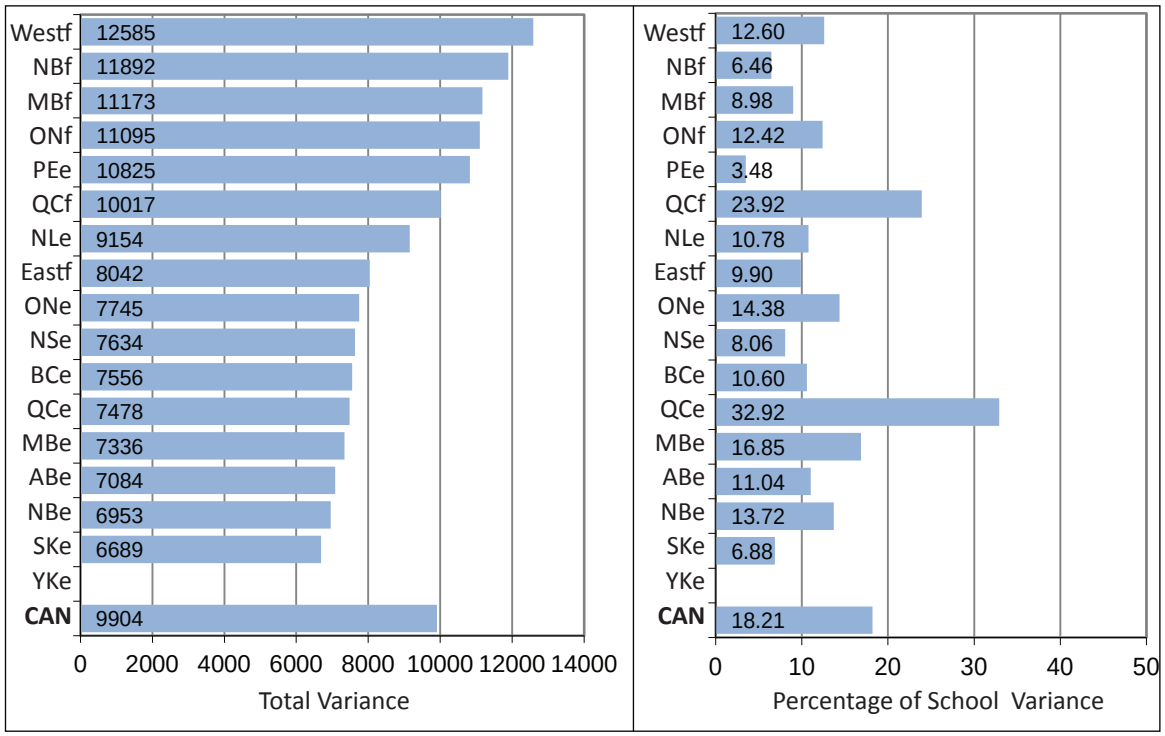
Providing "equality of opportunity" is a frequently stated goal of education systems, and is embodied in funding formulas, school programs, teacher allocations, and other jurisdictional policy instruments. The implicit assumption seems to be that equality of opportunity should result in equality of outcomes. However, this link is rarely examined. The concern of this chapter is with equity as indicated by the PCAP reading scores. Under the equity argument, the difference between the highest- and lowest-performing students on such indicators should be as small as possible. The equity principle can also be applied to schools, with the policy goal being to ensure that students in all schools perform at similar levels. Finally, the question of jurisdictional differences may also be expressed in equity terms, by stating that the goal is to have all jurisdictions within Canada perform at similar levels.

Reading Achievement Variance by Population

Equity may be examined in a number of ways. Following from the approach taken in developing the models, the overall variance for each of the jurisdiction/language groups is taken as a preliminary indicator of equity for both students and schools. On average, for Canada, the total reading score variance is approximately 10,000, as this is determined by the scaling of the scores to a standard deviation of 100 (the total variance is simply the square of the standard deviation). However, differences in the variance across jurisdictions/language groups may be interpreted as a broad measure of equality. The smaller the total variance, the more "equal" that system is in terms of the overall range of scores observed.

Chart 6.1 shows the total variance in reading scores for each population and the percentage of that total that is attributable to differences between schools. It is apparent from this that francophone populations generally have greater variation than anglophone populations, though Quebec French is close to the Canadian average, and East French is below. The overall variances may be interpreted as a measure of relative equality or inequality for student scores as a whole. Under this interpretation, francophone populations are generally less equal than the anglophone populations.

CHART 6.1 **Total Variance and Percentage School Variance**



YKe Missing: Model did not converge

On the percentage of school variance, the two Quebec populations stand out, showing much greater variation between schools than any other populations. Manitoba English is slightly above the Canadian average of 15% variance between schools, and Ontario English is close to that average, as expected given the high weight for that population in contributing to the average. Prince Edward Island has the least school variation, which likely relates to the small number of schools in that population. Saskatchewan English is notable for having both low overall variance and low between-school variance, even though its average achievement level is relatively low.

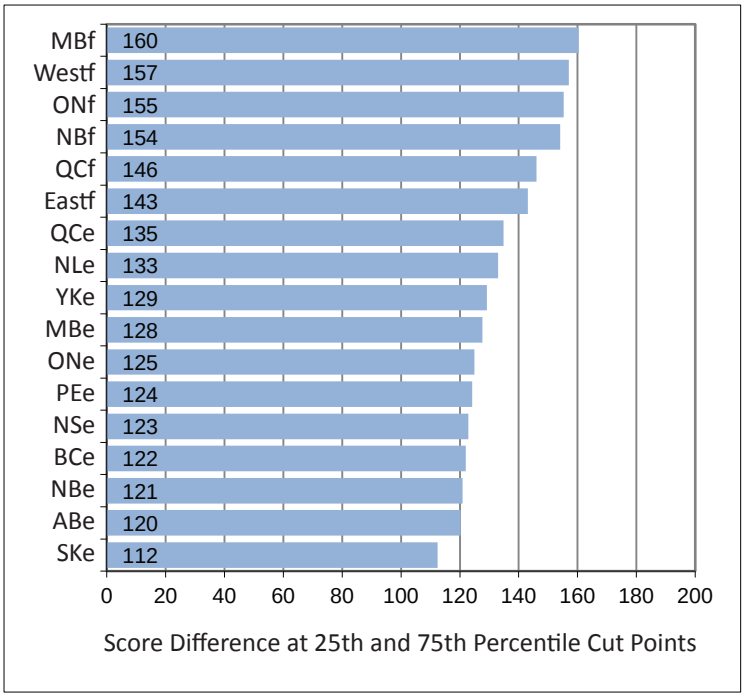
Reading Interquartile Range by Population

Another way of looking at equity is the approach taken by PISA, in which the score distributions are divided into quartiles (four groups of about equal size) and the cut score for each quartile computed. The "interquartile range," or the difference between the cut points for the 25th and 75th percentiles, may be used as an index of equity. This may be treated as a "reverse index" of equity, with higher numbers indicating less equity. The results for this approach are given in Chart 6.2. The pattern here is similar to that for the variances, with the francophone populations showing the largest differences.

By this measure, Saskatchewan English is the most equal jurisdiction, even though its average reading score of 471 is significantly below the Canadian average. The highest-performing population, Quebec French, is among the least equal by this measure. More generally, for Canada as a whole, the difference between the first and fourth quartiles is

approximately 124 points. All of the English populations are within ±10 points of this average, while all of the francophone populations are more than 10 points higher than this average. Generally, the francophone populations are more variable in both their average scores and their score distributions.

CHART 6.2 **Interquartile Ranges for Reading Scores by Population**

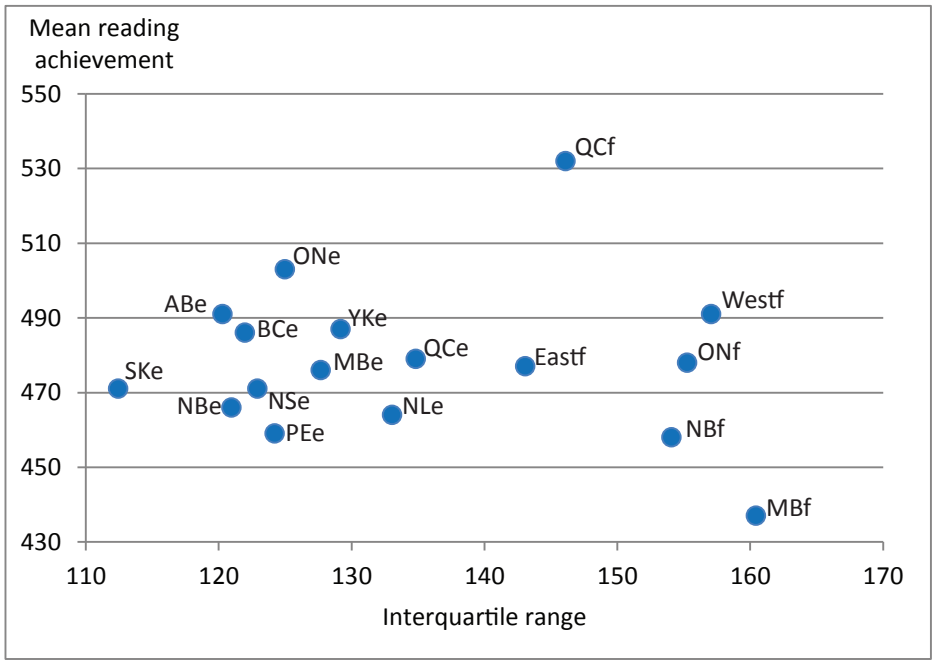


Equity and Achievement

The contrasting results for Quebec French and Saskatchewan English, in particular, raise the question of whether there is a trade-off between higher average achievement and greater equity in achievement. This point can be examined more closely by plotting the mean reading achievement score and the equity index on the same graph.

Chart 6.3 shows this plot. While making it clearer that all of the francophone populations have higher interquartile ranges (less equity) than any of the anglophone populations, this chart also indicates that the francophone populations also show a wider range of achievement. Anglophone populations tend to be more clustered at the low end of the interquartile range index and closer to the middle of the achievement distribution. Overall, the correlation between achievement and equity is close to zero.

CHART 6.3 **Mean Reading Achievement and Interquartile Range by Population**



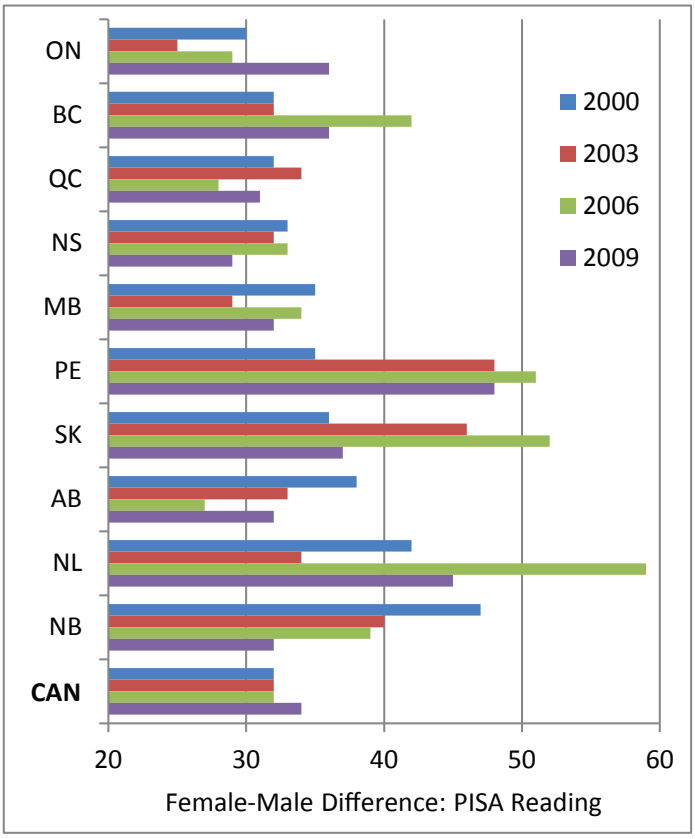
Gender Differences

Large-scale assessments have consistently shown that girls outperform boys in reading. Differences in mathematics and science have historically been much smaller and inconsistent. The reading gender gap has attracted considerable policy and public interest. One of the questions posed for this study is whether some jurisdictions have been more successful than others in reducing this gap.

The most comprehensive set of results on the reading gap is that provided by the PISA assessments. Four reading assessments, at three-year intervals from 2000 to 2009, are available for Canadian jurisdictions (no separate language breakdowns are available). Chart 6.4 gives these results.

Overall, for Canada, there has been little change in the size of the gender gap over the decade. For most jurisdictions, there has been some fluctuation but no distinct pattern. Ontario, British Columbia, Quebec, Nova Scotia, Manitoba, and Alberta have shown consistently smaller gaps than other jurisdictions. Large fluctuations are apparent for Saskatchewan and Newfoundland and Labrador. Only New Brunswick shows a consistent pattern of decrease in the magnitude of the gap.

CHART 6.4 **Gender Difference in PISA Reading Scores by Province, 2000-2009**[9]



Language Differences

The largest differences in equity in the PCAP-13 2007 reading results are clearly between the francophone and anglophone populations, taken as groups. A distinction between francophone and anglophone populations, and, in some cases, between Quebec French and other francophone populations, is also apparent in many of the results presented in the PCAP-13 2007 Contextual Report. This suggests that the remainder of the analysis focus on these two groupings. Some of the most striking distinctions, drawn from the relevant charts in that report, are given in Table 6.1.

While interesting in showing differences in student and school characteristics between the language groups, these results cannot be interpreted directly in terms of equity because they address the pattern of values for the variables but not the variation in these values within populations. The next section examines these variations more directly.

[9] Error bars are not plotted to reduce the complexity of the graph. Generally, confidence intervals for jurisdictions are in the range of ±8-10 points, and for Canada ±4 points.

TABLE 6.1 **Variables Showing Significant Differences between Anglophone and Francophone Populations, with Specific Reference to Quebec French**

Variable	Observed pattern	PCAP-13 2007 Contextual Report chart reference
Gender	Proportion of males less than 50% in most francophone populations	3.2
School enrolment	QCf tends to have larger schools than any other population	3.35
School governance	Both QCf and QCe have more private schools than other populations	3.38
External attributions of failure	Lower for francophones	4.13
Internal attributions of success and failure	Lower for francophones	4.16
Reading by decoding	Lower for francophones	5.3
Use of external sources in reading	Lower for francophones, especially for QCf	5.5
Academic/cultural activities	Higher for francophones	5.11
Emphasis on external assessments	Higher than average for most francophone populations	6.3
Class size	Largest for QCf by a wide margin	6.4
Student absenteeism	Generally lower for francophones for both students and schools	7.5, 7.7
Homework correction by individual students	Higher for francophones	7.17
Homework counts toward marks	Lower for francophones	7.17
Direct reading strategies	Higher for francophones	8.3
Indirect reading strategies	Lower for francophones	8.5
Use of creative reading materials	Lower for francophones	8.9
Teacher assignment of written reports	Lower for francophones	8.11
Use of media in reading	Lowest for QCf	8.15
Use of library/literature materials in reading	Lower for francophones	8.16
Assessment by short test items	Lowest for QCf	9.1
Assessment by long test items	Lower for all francophones except QCf	9.2
Assessment by non-academic criteria	Less use by francophones	9.13
Students know what a rubric is	Lower for francophones	9.16
Rubrics used for scoring	Lower for francophones	9.17
Teacher assessment skills	Higher for francophones	9.20

Factors Contributing to Equity

In principle, the interquartile range may be treated as a dependent variable and its effects modelled, as in the previous chapter. However, it is not technically feasible to use multi-level modelling in this case because the interquartile range is inherently a population-level variable, the value of which is constant for all students and all schools in a population. Any model for this index must therefore be restricted to the 17 units available at the population level. This is an inherent problem in the comparative analyses of education systems using system-level variables. There is no clear analytical solution to this problem. This leaves us with the ability only to do descriptive/comparative analyses, in which selected characteristics of the populations available are examined in relation to the outcome of interest, in this case the equity index.

The focus for the equity issue is not on mean scores but on the range of scores. Differences across jurisdictions in the equity index should thus be linked to differences in the distributions for variables that are correlated with the index. For example, if a variable such as "books in the home" is positively correlated with achievement, one would expect the amount of variation in books in the home to be correlated with the amount of variation in achievement for a population. Populations with a wider range of books in the home might be expected to have a wider range of achievement.

It is therefore worth looking at the distributions of some such variables. This is not an entirely straightforward matter because the variables of interest are on different scales, many of which are categorical in nature, which precludes computing interquartile ranges or similar indices. Also, it is difficult to present a concise summary of the distributions of a large number of variables over all populations. The task can be simplified by limiting the analysis to anglophone and francophone populations, where the largest differences in equity are found. The approach taken here is therefore to examine how reading scores change for the two language groups as a function of variations in the categories of a selection of predictor variables. Plots are given for the distributions of the selected predictor variables. Differences in the slopes for these plots show the "interaction" between these variables and language with respect to reading achievement. These are tested statistically by two-factor analysis of variance in each case.

Socioeconomic Status

Many equity analyses focus on the "socioeconomic gradient" as a key factor in equity. The essential argument is that, since socioeconomic status is a strong predictor of achievement, wider variations in socioeconomic status would be expected to contribute to wider variations in achievement. From a general social policy perspective, the goal might be to reduce the degree of socioeconomic disparity as a means of decreasing variation in achievement. From an educational policy perspective, however, the opposite might be the case. That is, the educational policy goal would be to use reduced achievement variation as a means of reducing socioeconomic disparity, at least intergenerationally.

This issue can be examined here by comparing the two available SES indices, books in the home and mother's education, for francophone and anglophone jurisdictions. This comparison is given in Charts 6.5 and 6.6.

It is apparent from these charts that the distributions for these two indicators are quite similar for English and French populations. The relationship between these indicators and reading achievement is also similar. However, the slope is slightly steeper for francophone than for anglophone populations. A test of the interaction between language and these two socioeconomic variables shows a statistically significant effect. Higher SES is associated with higher achievement, and this effect is more pronounced for francophones than for anglophones.

Although socioeconomic status is a consistent predictor of achievement and, it appears, of variation in achievement, this is not the area of most interest in this study. Instead, the interest is in variables that may be influenced by educational policy. Unfortunately, the bivariate approach to analysis that must be taken here limits the number of variables that can be discussed in any detail. A few variables judged to be of direct policy interest have therefore been selected for further discussion.

CHART 6.5 **Distribution of Socioeconomic Indicators by Language**

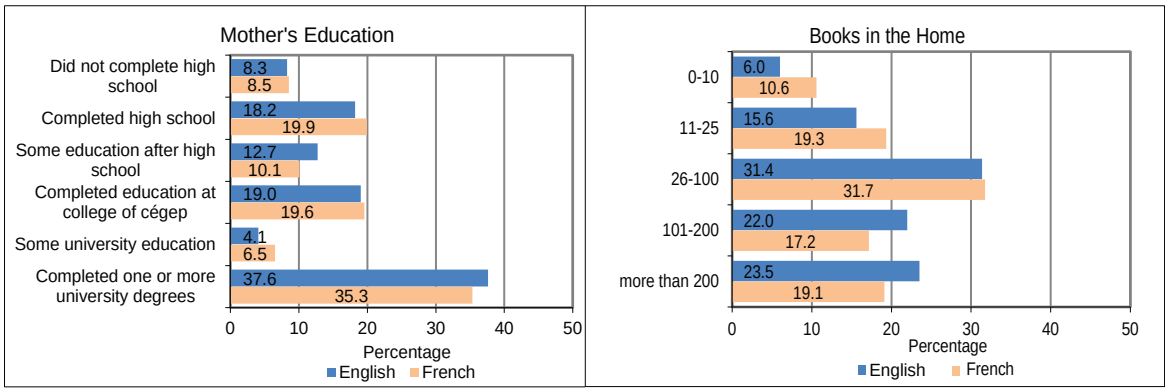
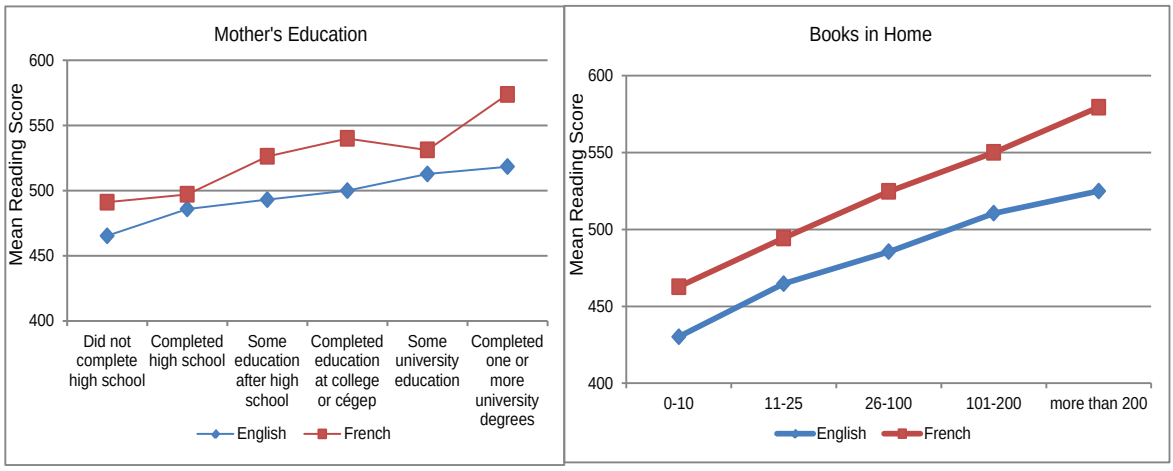


CHART 6.6 **Relationship Between Socioeconomic Indicators and Reading Scores by Language**



School Size and Community Size

Both school size and community size are positively related to achievement (Contextual Report, Charts 3.37 and 3.45). The distributions of these variables and the pattern of reading achievement for these variables by language are given in Charts 6.7 and 6.8.

These distributions show that francophone students are much more likely to be in larger schools and somewhat more likely to be in larger communities than anglophone students. Of course, these two variables are themselves correlated, as larger schools are more likely to be in larger communities. The achievement data show a significant interaction between these two variables and reading performance. Reading achievement is higher for students in larger schools and larger communities, and this effect is more pronounced for francophone than for anglophone students.

CHART 6.7 **Distribution of School Size and Community Size by Language**

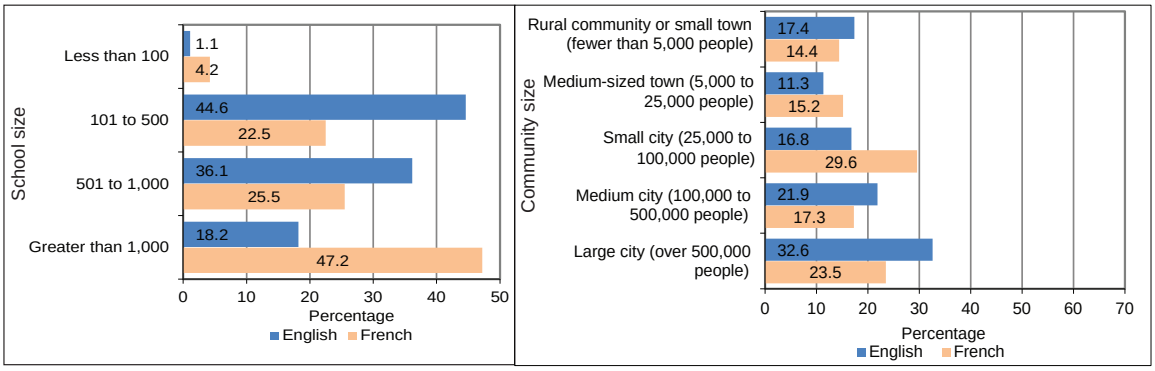
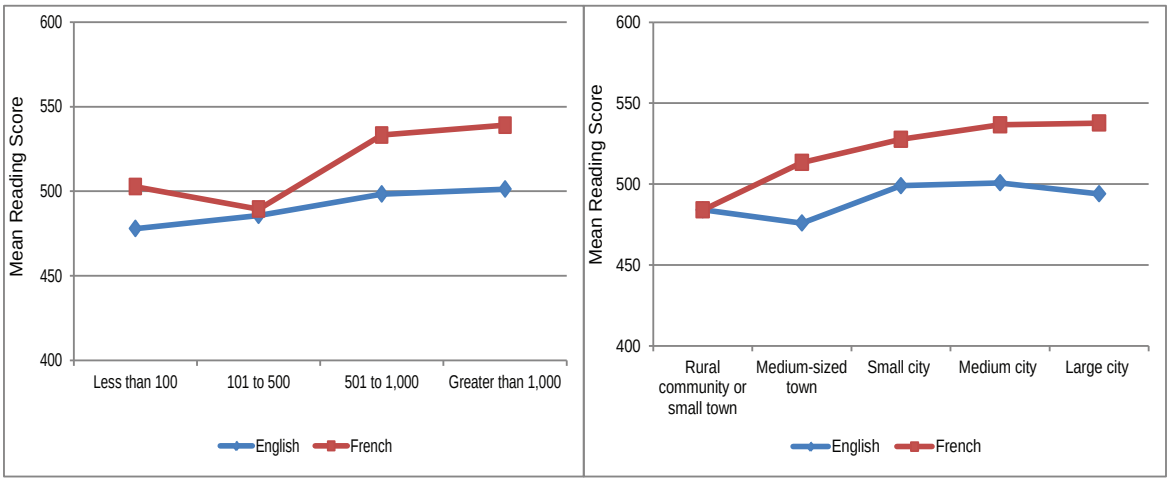


CHART 6.8 **Relationship Between School and Community Size and Reading Scores by Language**

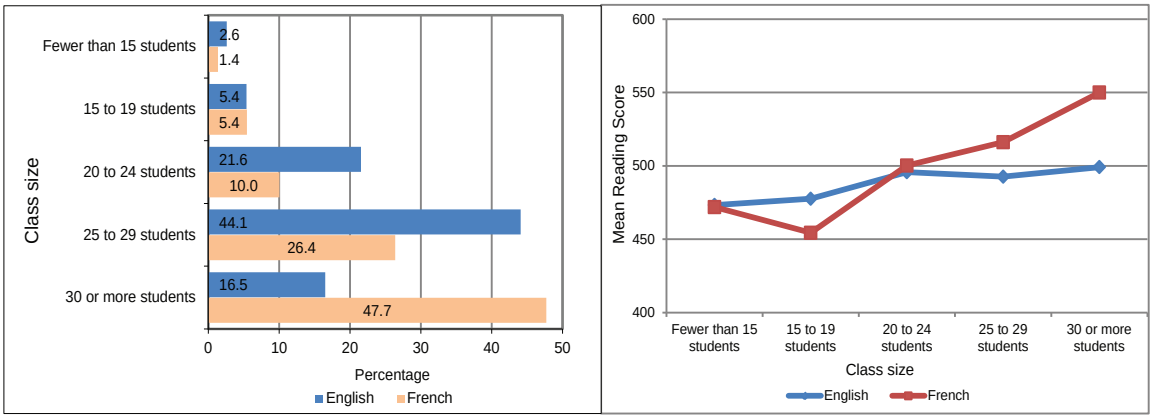


Class Size

The PCAP-13 2007 results show that, contrary to intuition and to most policy directions, students in larger classes have higher reading scores than those in smaller classes (Contextual Report, Chart 6.6).[10] This is true even when other variables (such as school and community size) that are correlated with class size, are controlled (Contextual Report, Chart 10.3).

Chart 6.9 shows the distribution of class size and the trend for class size and reading achievement by language. In this case, the distributions are quite different for the two language groups, with similar proportions at the low end and much higher proportions of large class sizes for francophones. The latter is actually almost entirely a phenomenon for Quebec French. Other francophone populations tend to have relatively small class sizes. The achievement graph shows that class size also affects achievement to a greater extent for francophone than for anglophone students, with this interaction effect being statistically significant.

CHART 6.9 **Distribution of Class Size and Relationship Between Class Size and Reading Achievement by Language**



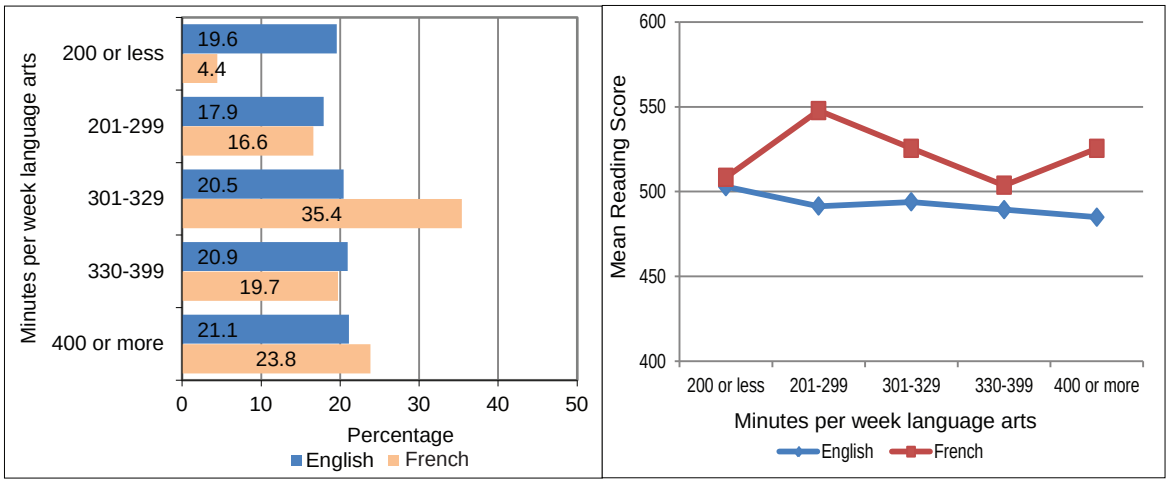
[10] This result is also inconsistent with recent experimental research on class size. However, most of the latter research involves students in the early grades and is not directly generalizable to the intermediate grades. Most large-scale assessments give results consistent with those found here.

Time on Language Arts

The Contextual Report (Chart 7.2) shows a negative bivariate relationship between time spent per week on language arts and reading achievement. This relationship largely disappears when other variables are controlled (Contextual Report, Chart 7.23).

Chart 6.10 shows the breakdown of minutes per week on language arts and the relationship with achievement by language. In this case, the distribution is broader for English than for French, with fewer francophone schools at the lower end of the distribution, and more in the middle range. The relationship with achievement is slightly negative for English but non-linear for French. Although the interaction effect is statistically significant because of the non-linearity, this pattern cannot be interpreted in any direct way in terms of equity.

CHART 6.10 **Distribution of Time on Language Arts and Relationship Between Time on Language Arts and Reading Achievement by Language**

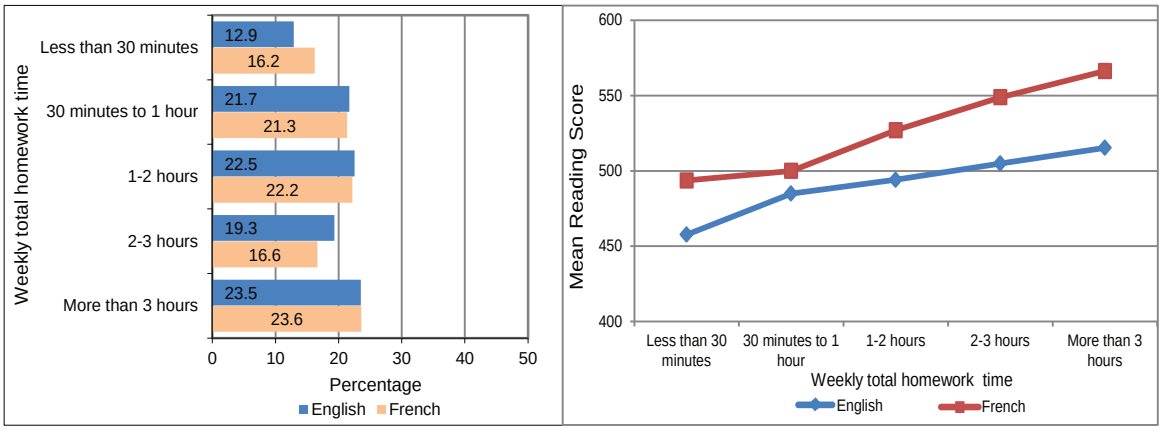


Homework Time

A positive relationship has generally been found between homework time and achievement (Cooper, Robinson, & Patall, 2006). In the PCAP-13 2007 Contextual Report (Chart 7.22), this was true for student-reported total time per week on homework. However, the relationship was less pronounced for time on language arts homework.

Chart 6.11 shows that the distribution of total homework time is quite similar for the two language groups. The interaction between homework time and language is significant, with amount of homework having a larger effect on francophone than on anglophone students.

CHART 6.11 **Distribution of Student-Reported Homework Time and Relationship Between Homework and Reading Achievement by Language**



Student Absenteeism

Two measures of student absenteeism are available in PCAP-13 2007 Contextual Report (Charts 7.6 and 7.9). Students reported their total days absent for the year, and principals reported average absentee rates for their schools. Both of these were found to be related to reading achievement, in the expected direction.

The distributions of these two indices by language are shown in Chart 6.12. In both cases, anglophones show higher absence rates than francophones. The interactions shown in Chart 6.13 are statistically significant, with the decrease in achievement with higher absence levels being greater for francophone than for anglophone students and schools.

CHART 6.12 **Distributions of Student and School Absence by Language**

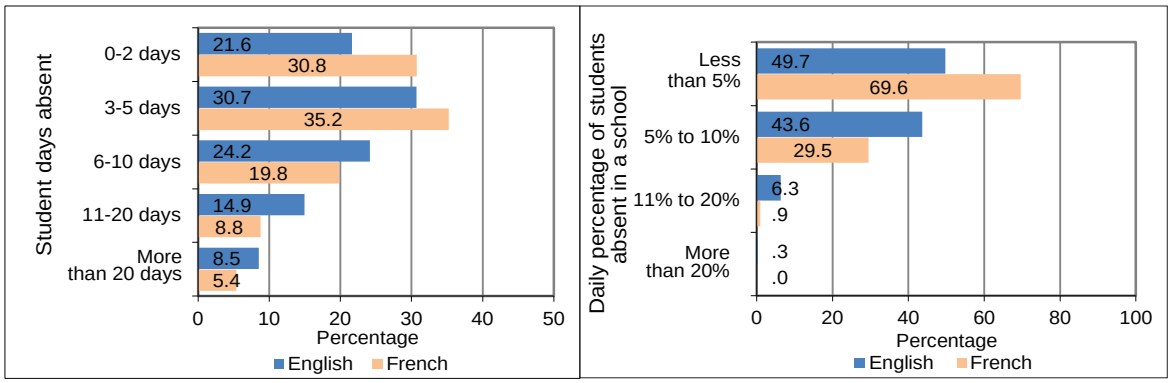
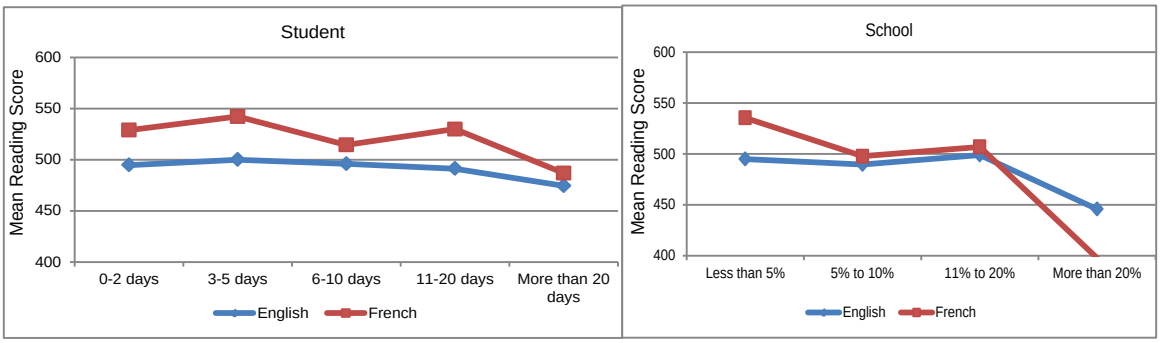


CHART 6.13 **Relationship Between Student and School Absence and Reading Achievement by Language**



In summary

The equity results present a consistent picture showing that variables known to be related to reading achievement exert a stronger effect on francophone than on anglophone students. The greater variability (or less equity) found in francophone jurisdictions seems to be less related to differences in the distributions of these variables than to the differential effects of these variables on achievement for the two language groups.

It is cautioned that the effects shown are bivariate in nature, with no controls for other variables that may be correlated with the ones examined. It is therefore not appropriate to draw strong causal conclusions from the trends found. Nevertheless, some of the factors examined here are amenable to change at the classroom or school level. In general, these are the ones that have to do with time to learn. For example, it is not particularly difficult to increase the amount of homework expected for those that are on the low end of the homework distribution. These results indicate that this can have a salutary effect on average achievement and, for francophones in particular, can help decrease achievement variability.

7 SUMMARY AND POLICY IMPLICATIONS

This report focuses on differences between jurisdictions and official-language groups within jurisdictions (collectively referred to as populations) in student achievement as measured by large-scale assessments and on equity in achievement within these populations. Following a historical review of results from the SAIP, PISA, and PCAP assessments from 1993 to 2009, a more detailed analysis of factors contributing to population differences and equity was conducted, using data from the PCAP-13 2007 assessment. This analysis was informed by an educational productivity model (Crocker, 2007), along with a literature review focusing mainly on research based on these large-scale assessments.

The Persistence of Jurisdictional Differences

The historical review reveals that the overall rankings of populations on these assessments have been remarkably stable over a period that encompasses close to two generations of students. Alberta and Quebec have consistently had the highest scores and the Atlantic provinces, the lowest scores. There are indications that British Columbia has had relatively better performance on the PISA assessments than on the others. The results also suggest that Ontario has shown some relative improvement over time. However, because of the large weight that Ontario contributes to the Canadian average, Ontario is consistently close to that average.

The magnitude of the differences on the SAIP assessments is of the order of 25%–30% of students meeting the expected standard (close to 80% of students meet the standard in the highest populations and only about 50% in the lowest). On PISA and PCAP, the magnitude is in the range of 0.60 to 0.70 standard deviation units. While it is not possible to make a definitive statement on whether there has been an overall change (decline or improvement) in achievement at a national level over the review period, the balance of evidence suggests that there has been little change.

Population Profiles

Based on the conceptual model and the literature review, variables showing statistically significant differences across populations and statistically significant effects on achievement were selected from the PCAP-13 2007 database. The primary reference for this selection was the PCAP-13 2007 Contextual Report (CMEC, 2009), which presents descriptive and comparative data on a large number of student, teacher, and school factors and their relationship to reading achievement.

As a first step in the analysis of population differences, a series of "population profiles" for students, teachers, and schools were developed from the selected variables. These reveal that even though all of the variables show significant differences across populations, the differences between adjacent populations in the distribution are generally small and are notable only when populations at the extremes of the achievement distribution are examined separately.

For the student-level variables, the following differences are notable:

- Both of the highest-performing populations (Ontario English and Quebec French) have more students in Grade 8 than the lowest-performing populations (Prince Edward Island English and New Brunswick French). Both of the latter actually have more students in Grade 9 than any other population. On average, including these lowest-performing populations, those in higher grades have higher reading scores. This indicates that the average score for these populations would have actually been lower if that population had the same proportion of Grade 8 students as the two highest-performing provinces.
- Similarly, both of the highest-performing populations have more students who are born outside of Canada compared to the lowest-performing populations. Since being born outside of Canada is negatively associated with reading performance, this should have the effect of lowering the scores for the higher performing populations. Obviously, this negative effect is being offset by other positive effects.
- Homework is positively associated with achievement, and students in the highest-performing populations tend to do more homework than those in the lowest-performing populations.
- For francophone populations, there is a difference in the number of books in the home. This difference is in the expected direction, relative to achievement, with Quebec (the highest-performing francophone population) having more books than New Brunswick (the lowest-performing francophone population).
- For the anglophone populations, more students in Ontario (the highest-performing English population) know what a scoring rubric is than is the case for Prince Edward Island (the lowest-performing English population). This is in the expected direction, since knowledge of rubrics is positively associated with achievement.

The main differences for the teacher variables are:

- Both of the highest-performing populations (Ontario English and Quebec French) have higher percentages of classes with more than 30 students than the low-performing populations. While this seems counterintuitive, this is consistent with the overall finding that students in larger classes have higher performance levels. This is true even after controlling for other variables that are related to class size (CMEC, 2009).
- Teachers in the two highest-performing populations expect more homework from their students than is the case for the low-performing populations.

- Teachers in Prince Edward Island English use homework for grading more often than those in Ontario English. There is no difference between the lowest- and the highest-performing francophone populations on this variable. This variable is negatively related to reading achievement.
- Quebec francophone teachers more often re-teach basic reading skills than their New Brunswick francophone counterparts. However, this variable is negatively related to achievement, and thus cannot help account for the higher achievement of Quebec francophone students. Again, this is likely a case in which the negative effects of this variable are offset by other positive effects.
- Higher assessment skill is positively related to achievement. Ontario English teachers reported having higher levels of assessment skill than Prince Edward Island English. However, there is no difference between the highest- and the lowest-performing francophone populations on this variable.
- Using non-academic criteria for grading is negatively related to achievement. This occurs much more often for Prince Edward Island English teachers than for either their Ontario English counterparts or those in either of the francophone populations.

Finally, the school variables show the following differences between the highest- and the lowest-achieving populations:

- More schools in Ontario English and Quebec French, the high-performing populations, are larger and are in larger communities than in the low-performing populations.
- In Quebec French, there are more private schools than in other populations. Students in private schools, especially in Quebec, have higher reading achievement than those in public schools, even after controlling for other variables (CMEC, 2009). The private school effect is thus a plausible factor in accounting for the exceptionally high performance of Quebec francophone students.
- New Brunswick francophone schools (the lowest-performing francophone population) are more likely to spend more than 300 minutes per week on language arts than is the case for other populations. This variable is negatively related to achievement, which suggests that schools in which students are not performing well in language spend more of their time on that subject without this having the desired effect on achievement. The problem here, of course, is that there is no way to tell whether spending more time has a positive effect on those who are given the added time. What is clear is that the added time is not having a decisive effect in overcoming lower average achievement.

Modelling Achievement Differences

The starting model for this phase of the analysis was a "population model" in which each population was coded 0 or 1 as a separate variable in the model. These codes were used as independent variables in a two-level (student and school) hierarchical model with PCAP-13 2007 reading scores as the dependent variable. The coefficients generated by this model represent the difference between the reading score for each population and a "reference population."

The specific interest here is in the changes in these coefficients as various clusters of independent variables, drawn from the PCAP-13 2007 Contextual Report, were added to the model. The change in a population coefficient, as variable clusters were added to the model, was taken as a measure of the effect of the particular cluster (and of individual variables within the cluster) on the difference between that population and the reference group. The main part of the analysis focused on a "full model," which included all of the variables identified. A statistically significant difference between a population coefficient in the full model and in the initial population model was taken to indicate that the complete set of independent variables has an effect on the difference between a population and the reference group.

This analysis revealed that, with only one exception (New Brunswick French), the full model coefficients were not significantly different from the population model coefficients. The general conclusion is therefore that the set of variables used in the model do not account broadly for population differences. Nevertheless, examination of the intermediate models (Table A.1) reveals a few notable effects.

- For New Brunswick French, a statistically significant shift in the coefficient is found when student demographic variables are added to the model. This shows that controlling for student demographics (gender, grade, born outside of Canada, language match, and number of books in the home) reduces the difference between New Brunswick French and the reference group, Ontario English.
- Within the student demographic variables, the largest effects on achievement for New Brunswick French are for grade and language match. However, these variables cannot account for the achievement difference between that population and Ontario English because they have similar values for both populations. The greatest contrast between the two is for books in the home. Ontario English students have, on average, many more books than New Brunswick French students. The latter variable is our best measure of socioeconomic status, thus suggesting that part of the shift in coefficients for New Brunswick French is related to socioeconomic status.
- The coefficient for New Brunswick French begins to increase again as school characteristics, teacher characteristics, and student reading strategies are added to the model. This suggests that some of the variables in this category are having a negative effect on the performance of this population. None of these variables has a particularly strong effect on achievement, and some have negative and others positive effects. It is therefore difficult to attribute the observed change to any specific variables within these clusters. The coefficients actually decrease again (become less negative) as the

remaining clusters are added to the model, indicating that instructional climate, time, teaching strategies, and assessment strategies and practices have overall positive effects on New Brunswick French students.

A few other statistically significant changes for other populations are found in the intermediate models. These do not show up in the full model because of various complex interaction effects that were not analyzed. Examples are:

- For Saskatchewan English, the coefficient becomes significantly less negative when school characteristics are added to the model. The most likely source of this shift is that adding this cluster controls for the relatively large number of Saskatchewan English schools with a high proportion of Aboriginal students.
- School characteristics also show a significant positive effect for Ontario English. Since this shift is relative to Quebec French, the most likely source of the shift is that this is a consequence of controlling for the large positive private school effect for Quebec French.
- For Quebec French, adding school characteristics significantly reduces the coefficient. The most obvious source of this shift is the large private school effect for this population, compared to that for Ontario English. However, other positive school characteristics, including school size and class size, are also favourable to the Quebec French population.

Achievement Equity

The equity issue relates not to differences across populations but to variations in achievement within a population. The policy goal here is that jurisdictions should strive not only toward high average achievement but also toward reducing the disparities between those at the high and low ends of the achievement distribution. More specifically, achieving greater equity is often thought of in terms of increasing the performance of those at the lowest end of the distribution. The literature indicates that this can be done without diminishing the already high achievement of those at the top. Improvement at the low end can also translate into an increase in average achievement if the improvement occurs in sufficient numbers of students.

Equity was examined first by looking at the total variance of the PCAP-13 2007 reading scores for each population. This variance can be divided into two components, between students and between schools. The latter is of most interest because difference between schools is a matter of direct policy concern.

On average for Canada, on the PCAP scale, the total variance has a value of 10,000 (the square of the standard deviation). Across populations the variance ranges from about 6,700 to about 12,500, or -33% to +25% of the Canadian average. The between-school variance shows even greater divergence, ranging from about 3% to 40% of the total variance. Francophone populations generally show greater variance than anglophone populations.

As a second measure of equity, the interquartile range (the range between the 25th and 75th percentiles on the reading scale) was computed for each population. The picture for this measure is essentially the same as that for the variance. The interquartile range varies from 112 points to 160 points on the scale. A plot of equity versus achievement confirms that francophone populations show less equity than anglophone populations but also shows that there is essentially no relationship between average achievement and equity. Francophone populations generally have lower equity and are widely dispersed on the achievement scale. Anglophone populations are more clustered, at relatively high equity, and closer to the middle of the achievement scale.

Because the strongest contrast in equity is between the two language groups, the remainder of the analysis focused on these groups. A number of variables that show consistent differences between the language groups were identified from the PCAP-13 2007 Contextual Report (CMEC, 2009). These include socioeconomic status, school and community size, class size, weekly class time on language arts, homework time, and student absenteeism. These comparisons present a consistent picture showing that variables related to reading achievement exert a stronger effect on francophone than on anglophone students. The greater variability (or less equity) found in francophone populations seems to be less related to differences in the distributions of these variables than to the differential effects of these variables on achievement for the two language groups.

Responses to Research Questions

What does the research literature, particularly that based on large-scale assessments, tell us about factors that contribute to jurisdictional differences in achievement?

The research literature tells us a great deal about factors that contribute to achievement but less about factors that contribute to jurisdictional differences in achievement. This may seem contradictory because, presumably, the same factors that contribute to achievement should also contribute to jurisdictional differences. The problem is that most of the research does not focus explicitly on jurisdictional differences and that which does is plagued by methodological difficulties. These difficulties stem mainly from the relatively small number of units available for analysis at the jurisdictional level relative to the student and school levels. Specifically, the amount of variation between jurisdictions, compared to that between students and schools, is so small that it makes it virtually impossible to develop comprehensive statistical models with jurisdictions as the main independent variable.

The summary of the international comparative studies emphasizing jurisdictional differences that was given at the end of Chapter 2 is repeated here.

- Many factors influence school achievement. None has a decisive effect, and the cumulative effects of all factors included in the various studies have not been well established.

- Differences within jurisdictions are more important than differences between jurisdictions.
- Broad policies and reforms at the jurisdictional level can be important, as evidenced by improvements in achievement in some countries that have instituted major reforms.
- Absolute resource levels seem not to matter as much as how the resources are used. Within a given range, increasing resource levels seems to make little difference to achievement.
- Socioeconomic status is important. Almost all low-achieving countries have relatively low average SES. However, not all high SES jurisdictions have high achievement. The effects of socioeconomic status can be mitigated by policies designed to promote equity.
- High achievement and a high degree of equity are not conflicting goals. Some countries, including Canada, have achieved both high performance levels and a relatively high degree of equity among students.
- Countries with highly differentiated schools, through tracking or streaming of students, have variable achievement but less equity in achievement.
- Internationally, having a first language other than the language of the school tends to have a negative effect on achievement.
- Gender differences strongly favour girls in reading. The size of the gender gap varies widely across countries, suggesting that reading achievement can be improved in some jurisdictions by focusing on improving the performance of boys.
- There are indications that systems with more frequent testing and external exit exams tend to have higher achievement.
- The high level of both achievement and equity in Finland has been attributed to the existence of a comprehensive network of schools and the recruitment of highly qualified teachers in all schools.
- In terms of economic growth in highly developed countries, achievement of basic skills has more payoff than achievement of high skills and tertiary schooling in the long run.
- The overall performance of Canada has declined slightly over the past decade. Nevertheless, Canada remains one of the highest-performing countries, while combining this with a high degree of equity.

Do some jurisdictions attain greater equity (lower variation) than others in achievement?

The short answer to this question is "yes," at least for PCAP-13 2007 reading. A more precise answer is found in the language breakdowns. Specifically, for PCAP-13 2007 reading, anglophone populations show less variation in achievement than francophone populations. Although many of the predictor variables used in this analysis have similar

distributions across anglophone and francophone populations, the evidence given suggests that these predictors have a greater effect on achievement in francophone than in anglophone jurisdictions.

The PCAP-13 2007 results also support the conclusion from international studies that there is essentially no relationship between equity and achievement. Some countries, Canada included, have performed relatively well by both measures.

Were some jurisdictions able to narrow the gap between sub-populations (e.g., boys/girls, francophones/anglophones)? Is this related to or independent of average achievement?

Girls continue to outperform boys in reading. The PISA assessments offer the only consistent way to examine this gap over time. For most Canadian jurisdictions, the size of the gender gap has fluctuated over the four assessments available. However, only New Brunswick shows a pattern in which the gap has consistently declined over time. In the absence of consistent patterns, it is not possible to say anything about whether the gap is related to or independent of average achievement.

What combinations of factors contribute to particularly high or low achievement within and across jurisdictions?

The evidence presented here, and in the many reports from SAIP, PISA, and PCAP, points to many factors that are related to achievement. The general pattern shows that the effects of various factors tend to be consistent, at least in direction, across jurisdictions. While most effects are relatively small, their consistency provides strong evidence that these are universal and are not artefacts of particular settings. Most of the effects are consistent with the Wang/Haertel/Walberg proximity model or with the Carroll time model. However, some time variables, such as length of the school year or day, that are most amenable to policy change but do not vary much across Canadian populations, cannot be expected to show differences across jurisdictions. Other student-level time variables such as reading outside of school, homework, and absenteeism do show small but consistent effects.

While this question embodies the core issue of concern in this study, that of differences across jurisdictions, the study has revealed that it is difficult to investigate this issue using data from large-scale national assessments. This stems largely from low levels of interjurisdictional variance in both achievement and factors affecting achievement. This yields little that can be modelled using standard statistical techniques that are predicated on having sufficient “real” variation to offset the inherent variation due to sampling and measurement error.

Are there particular combinations of school, teacher, and/or student characteristics that may offset the well-established effects of socioeconomic status on achievement?

In the model used in this study, and in others reported in the PCAP-13 2007 Contextual Report, the effect of socioeconomic status is attenuated as more variables are added to the model. For example, in the PCAP-13 2007 Contextual Report, the effects of both mother's education and books in the home are significantly reduced in the full model compared to the bivariate model. Nevertheless, both remain significant predictors of achievement. In the model used in this report, the effect of books in the home increases when the populations are added to the model, suggesting that this effect is suppressed by population-specific effects. However, in the full model, the effect is again reduced, to about the same level as in the bivariate model.

Policy Issues

This study rests on the assumption that the differences in achievement between jurisdictions and language groups (collectively referred to as populations) observed in large-scale pan-Canadian and international assessments are large enough to attract policy attention. The historical results reinforce this by showing that these differences have been persistent, with the rankings for these populations having changed very little since the early 1990s. In a country with many institutions and arrangements devoted to equality among its main political and cultural groups, a case can be made that a high degree of equality of educational achievement is essential to ensure that other forms of economic and social equality can be achieved.

The same point can be made about equity among students and schools within populations. Achievement differences between schools within a population also raise significant policy concerns within the provinces and territories because such differences indicate that where a person goes to school can have an important impact on that person's ability to maximize educational opportunities.

Achievement differences between students is a rather more complex matter because of the large number of factors that can contribute to such differences, many of which are outside the control of the school system. Nevertheless, minimizing differences between students (thus improving equity) is a well-established goal of education. In particular, there are significant policy thrusts in the direction of improving the performance of the most disadvantaged students. If this can be accomplished without sacrificing the performance of high-achieving students, both average achievement and equity can be improved. If the spread of achievement is larger in some jurisdictions than in others, it is important to determine if this is related to features that can be influenced by policy.

This study shows that the variables available in PCAP-13 2007 account for only a very small proportion of the variation between populations, once statistical error is taken into account. The ability to detect "real" differences across populations depends on having sufficient sample size, sufficient variation across the populations being analyzed, and sufficient correlation between the outcomes of interest and the factors related to those

outcomes. None of these conditions are well met in the available databases. In simple bivariate analyses, the effects of one variable are confounded with those of many other variables. In complex modelling designs, the observed effects are often too small to be detected.

Beyond the factors identified in the PCAP data, are there other characteristics of school systems in different jurisdictions that may contribute to differences in achievement levels?

Aside from statistical issues, it is possible that large-scale assessments are simply not capturing the right variables. There is a reasonable possibility that population differences are caused by large-scale structural features of the system that are not measured by these assessments. Over the period examined in this study, many jurisdictions in Canada have undergone significant structural changes and other substantive reforms. This was especially true in the 1990s when Royal Commissions and other major public inquiries were completed in a number of jurisdictions. A few examples are:

- In two jurisdictions, Quebec and Newfoundland and Labrador, these have resulted in major changes in governance, including constitutional change, with the rights of religious denominations being significantly diminished.
- Several jurisdictions have introduced or enhanced large-scale assessments. The most notable example of this is in Ontario, where an arm's-length body, the Education Quality and Accountability Office (EQAO) has been established. EQAO has implemented a number of province-wide assessments, including a Grade 10 Literacy Test, which students must pass to graduate from high school.
- New Brunswick has implemented a large-scale shift to a middle-school structure, with separate schools for Grades 6 to 8 and Grade 9 becoming part of the high school system. New Brunswick has also substantially altered the role of school boards in governance.
- Almost all jurisdictions were faced with serious financial austerity in the 1990s, leading to consolidation of school boards and schools.
- The general trend in many jurisdictions is toward greater centralization of the curriculum and greater emphasis on outcomes.
- The inclusion principle for special-needs students has come into full force in most jurisdictions, with almost all special-needs students now being integrated into regular schools and classrooms.

Of course, not all such structural changes are intended to improve achievement in core subjects in the intermediate grades. Change may be driven by a host of other goals such as efficiency or cost control, human rights, access issues, advocacy or political action, union demands, or simply the beliefs of those with the power to bring about change. Indeed, the Wang/Haertel/Walberg model suggests that large-scale structural features of the system are the least likely to influence achievement. Even though these features cannot be explicitly

captured by the type of models used here, the lack of significant shifts in achievement on the part of most populations suggests that population-level innovations of the type given above are not significant causes of achievement change; otherwise, at least some jurisdictions should have shown significant shifts over the time frame under review.

Other than structural features, are there variables not being captured by large-scale assessments that may have a significant influence on achievement?

Several possibilities may be identified. The first is curriculum content. Although curriculum development has evolved into a sophisticated enterprise, with ever greater emphasis on outcomes, the effects of curriculum change are largely unknown. While most curriculum changes begin with "pilot projects" of new programs, these are typically concerned with the mechanics of implementation and with such factors as teacher attitudes toward the change. Rarely is curriculum change systematically investigated for its impact on outcomes. It is even rarer to find large-scale randomized clinical trials of new programs. There seems to be an implicit assumption that a new program will be an improvement on the one it replaces.

A second, and related, issue is curriculum implementation. What is actually done with the curriculum in schools and classrooms, and hence the content to which students are actually exposed, is rarely investigated. Some large-scale studies have developed measures of "opportunity to learn." However, these are difficult to implement through questionnaire methods. The gap between the intended curriculum (on which many large-scale assessments, including PCAP, are based) and the implemented curriculum may simply be larger in some jurisdictions than in others.

The same point may be applied to assessment practices and particularly to the expectations that may be built into scoring and grading. Since advancement through the school system, including high school graduation, is largely based on teacher judgments, there is no assurance that similar grades represent similar performance levels across schools or systems. As long as students appear to be performing satisfactorily within their own systems, there is little reason to be much concerned with performance on external, low-stakes assessments such as PISA or PCAP or even provincial assessments in most cases.

The latter point raises the issue of expectations. Although large-scale assessments have been around for about two decades, there is little to indicate that the results of these assessments have a very high profile outside of provincial/territorial education departments. For example, media reports of large-scale assessments often focus on PISA and specifically on the fact that Canadian students do quite well in an international setting. Rarely are the jurisdictional differences in these results examined in any detail. If teachers, principals, parents, and the general public are largely unaware of the results of these assessments, there is little to drive any action toward improvement. Departments of education alone cannot improve achievement unless there is a clear sense at the local level that it needs to be improved. Unless improving achievement is an explicit goal, shared at all levels of the system, it is difficult to see how improvement can occur.

One possibility is that public expectations lie more in the direction of "attainment" than achievement. That is, the concern may be more with "making it through the school system" than on what is learned along the way. To illustrate this point, the problem of dropping out of high school was a matter of much public concern throughout much of the latter part of the 20th century. This concern has largely dissipated in most jurisdictions (except perhaps for specific high-risk groups) in the face of dramatic improvements in the number of students graduating from high school. As long as students can complete schooling (with large numbers having access to postsecondary education), it is possible that achievement is not much of an issue, even in jurisdictions where achievement levels are relatively low. If, as the data suggest, we have significantly increased the number of high school graduates, without diminishing average achievement, then this can be considered a change in the right direction.

Although no single factor will be decisive in any efforts to improve achievement, and some of the most influential factors are outside the control of the school system, a good deal is known about structures that promote high achievement, and about teaching and learning strategies that can be effective. Even if these cannot account for jurisdictional differences, this does not change their overall impact on achievement. It is also known that simply increasing resource allocations is not likely to be effective unless these are directed toward enhancing the factors that are known to have positive effects.

Any jurisdiction wishing to embark on a systematic thrust to improve achievement can use some of the results of large-scale assessments to bring about changes that would increase the probability of success. Many such changes can be implemented without significant additional resource outlays. While departments of education can exercise leadership, this can be effective only if there is consensus among all stakeholders that something needs to be done. At this point, it is doubtful that such a consensus exists or that most stakeholders are actually aware that students in their jurisdictions are not doing as well as those in other jurisdictions within Canada.

Research Design Issues

This study has identified some significant analytical issues in research on jurisdictional differences. However, it may also be argued that large-scale assessments, as currently designed, are not optimal for the study of cross-jurisdictional effects. To begin with, many of the variables that have been found to influence achievement act in much the same way in all jurisdictions. Indeed, it would be unusual if this were not so. While the adverse effects of external variables such as socioeconomic status may be mitigated by jurisdictional policies, there are no instances in which the direction of these effects is reversed. The same is true for the teaching and learning variables that are typically measured in large-scale assessments. It is reasonable to expect, for example, that if more time improves learning, this will be so in all jurisdictions. Differences between jurisdictions on such factors are relatively small, especially in Canada where, even in the absence of any significant national educational policies, it is reasonable to conclude that jurisdictions are more alike than different on many of the factors affecting achievement.

This leads to the question of whether different approaches to research could be employed to help shed more light on jurisdictional differences. The point that we may simply be looking at the wrong factors has been made above. Studies that are more explicitly focused on significant reforms to structures, resources, curriculum, treatment of special-needs students, and the like are clearly needed. Such studies would have to be different from current large-scale assessment designs in at least two ways. First, they need to be immediate and focused on specific innovations. It is unreasonable to expect, for example, that the impact of a policy directed at improving early childhood literacy must wait until students are in Grade 8 or 10 to be measured in PCAP or PISA or that these effects could be disentangled from many other intervening events over that period of time.

Research on specific innovations needs to be conducted in the immediate context of these innovations. For example, the current practice of introducing curriculum change through pilot studies followed by full-scale implementation does not constitute an adequate research design. While we have argued elsewhere for the need for experimental studies (Crocker, 2008), the prospects of a significant shift to experimental research designs remain slim. Nevertheless, quasi-experimental designs such as time series and comparative designs are possible if properly focused on specific innovations.

Related to this, it takes a long time for the effects of many large-scale policy changes to take effect. Curriculum change is again a good example of this, as are policies intended to accommodate special-needs students, resource increases or resource reallocations. A change to the elementary school language arts curriculum, for example, is typically phased in over several years, working upward through the grades. Although the full impact of the change might not be manifest for several years, it is important to track the change from the beginning to reduce the impact of confounding variables and to design studies that would allow some control over such variables as well as tracking the impact of the change in schools and classrooms. Indeed, one of the important issues in curriculum change and similar innovations is that their intent may be lost in implementation. All of this suggests the need for longitudinal studies, focusing on specific innovations that have occurred in some jurisdictions and not in others. Such studies would not only monitor achievement effects more closely than is possible with current cross-sectional studies, but could also shed light on why so many innovations seem to have no effect on performance differences across jurisdictions.

REFERENCES

Becker, G.S. (1993). *Human capital: a theoretical and empirical analysis, with special reference to education*. (3rd ed.). Chicago: University of Chicago Press.

Bishop, J. (1997). The effect of national standards and curriculum-based exams on achievement. *American Economic Review, 88*(1), 266–75.

Bussière, P., Cartwright, F., Crocker, R., Ma, X., Oderkirk, J., & Zhang, Y. (2001). *Measuring up: The performance of Canada's youth in reading, mathematics and science. OECD PISA study — First results for Canadians aged 15.* Catalogue no. 81-590-XIE. Ottawa: Statistics Canada.

Bussière, P., Cartwright, F., Knighton, T., & Rogers, T. (2004). *Measuring up: Canadian results of the OECD PISA study. The performance of Canada's youth in mathematics, reading, science and problem solving. 2003 first findings for Canadians aged 15*. Catalogue no. 81-590-XPE — No. 2. Ottawa: Statistics Canada.

Bussière, P., Knighton,T., & Pennock, D. (2007). *Measuring up: Canadian results of the OECD PISA study. The performance of Canada's youth in science, reading and mathematics. 2006 first results for Canadians aged 15*. Catalogue no. 81-590-XPE — No. 3. Ottawa: Statistics Canada.

Carroll, J.B. (1963). A model for school learning. *Teachers College Record, 64*(8), 723–33.

Carroll, J.B. (1989). The Carroll model: A 25-year retrospective and prospective view. *Educational Researcher, 18*(1), 26–31.

Cooper, H., Robinson, J.C., & Patall, E.A. (2006). Does homework improve academic achievement? A synthesis of research, 1987–2003. *Review of Educational Research*, 76(1), 1–62.

Coulson, A. (2009). Comparing public, private and market schools: The international experience. *Journal of School Choice, 3,* 31–54.

Council of Ministers of Education, Canada. (1997). *SAIP Science. School Achievement Indicators Program 1996.* Toronto: Author.

Council of Ministers of Education, Canada. (1998). *SAIP Mathematics. School Achievement Indicators Program 1997*. Toronto: Author.

Council of Ministers of Education, Canada. (1999). *SAIP Reading and Writing. School Achievement Indicators Program 1998*. Toronto: Author.

Council of Ministers of Education, Canada. (2000). *Report on Science II Assessment. School Achievement Indicators Program 1999*. Toronto: Author.

Council of Ministers of Education, Canada. (2003). *Mathematics learning: The Canadian context. School Achievement Indicators Program Mathematics III 2001*. Toronto: Author.

Council of Ministers of Education, Canada. (2003). *Student writing: The Canadian context. School Achievement Indicators Program Writing III 2002*. Toronto: Author.

Council of Ministers of Education, Canada. (2005). *Report on Science III Assessment. School Achievement Indicators Program 2004*. Toronto: Author.

Council of Ministers of Education, Canada. (2008). *PCAP-13 2007: Report on the assessment of 13-year-olds in reading, mathematics, and science*. Toronto: Author.

Council of Ministers of Education, Canada. (2009). *PCAP-13 2007: Contextual report on student achievement in reading*. Toronto: Author.

Council of Ministers of Education, Canada. (2010). *PCAP-13 2007: Report on differences in reading performance of 13-year-olds based on language and minority majority language status*. Toronto: Author.

Council of Ministers of Education, Canada. (2011). *Factors contributing to performance in PCAP 2007 mathematics and science*. Toronto: Author.

Crocker, R. (2007). Human capital development and education. In Piccott, G., Saunders, R., & Sweetman, A. (Eds.). *Perspectives on human capital development*. Montreal: McGill-Queen's University Press.

Crocker, R. (2008). *Credible and useful evidence for education*. Alberta Initiative for School Improvement Colloquium, October 20-22, 2008. (http://education.alberta.ca/media/6451036/credible_and_useful_evidence_for_education-bob_crocker.pdf)

Finn, J.D., & Achilles, C.M. (1999). Tennessee's class size study: Findings, implications, misconceptions. *Educational Evaluation and Policy Analysis, 21* (2), 97–109.

Fuchs, T., & Wöessmann, L. (2004). *What accounts for international differences in student performance? A re-examination using PISA data*. Discussion Paper 1287. Bonn, Germany: Institute for the Study of Labour.

Grissmer, D., Flanagan, A., Kawata, J., & Williamson, S. (2000). *Improving student achievement: What state NAEP scores tell us*. Arlington, VA: RAND.

Gunnarsson, V., Orazem, P., Sanchez, M., & Verdisco, A. (2009). Does local school control raise student outcomes?: Evidence on the roles of school autonomy and parental participation. *Economic Development and Cultural Change, 58*(1), 25–52.

Haahr, J.H., Nielsen, T.K., Hansen, M.E., & Jakobsen, S.T. (2005). *Explaining student performance. Evidence from the international PISA, TIMSS and PIRLS surveys*. Danish Technological Institute. Retrieved from http://ec.europa.eu/education/pdf/doc282_en.pdf

Hanushek, E., & Wöessmann, L. (2010). *How much do educational outcomes matter in OECD countries?* Retrieved from http://www.ifo.de/portal/page/portal/DocBase_Content/WP/WP-CESifo_Working_Papers/wp-cesifo-2010/wp-cesifo-2010-11/cesifo1_wp3238.pdf

Jürges, H., & Schneider, K. (2004). International differences in student achievement: An economic perspective. *German Economic Review, 5*(3), 357–80.

Jürges, H., Schneider, K., & Büchel, F. (2003). *The effect of central exit examinations on student achievement: Quasi-experimental results from TIMSS Germany.* CESifo Working Paper No. 939.

Knighton,T., Brochu, P. & Gluszynski, T. (2010). *Measuring up: Canadian Results of the OECD PISA Study The Performance of Canada's Youth in Reading, Mathematics and Science: 2009 First Results for Canadians Aged 15*. Catalogue no. 81-590-XPE — No. 4. Ottawa: Statistics Canada.

Ma, X. & Crocker, R. (2007). Provincial Differences in Reading Achievement. *Alberta Journal of Educational Research, 53 (1),87-109.*

Marzano, R. J. (2003). What works in schools: Translating research into action. Alexandria, VA: Association for Supervision and Curriculum Development.

Mincer, J., (1974). *Schooling, Experience and Earnings.* New York: Columbia University Press.

Molnar, A., Smith, P., & Zahorik, J. (2000). "1999-2000 Evaluation results of the students achievement guarantee in education (SAGE) program." Milwaukee, WI: School of Education, University of Wisconsin-Milwaukee. Retrieved from http://nepc.colorado.edu/files/cerai-00-34.pdf.

Nye, B., Hedges, L. V., & Konstantopolous, S. (1999). The long-term effects of small classes: A five-year follow-up of the Tennessee class size experiment. *Educational Evaluation and Policy Analysis, Special Issue, 21* (2), S127–142.

Organisation for Economic Co-operation and Development. (2004). *Learning for tomorrow's world: First results from PISA 2003*. Paris: Author.

Organisation for Economic Co-operation and Development. (2007). *Learning for tomorrow's world: First results from PISA 2006*. Paris: Author.

Organization for Economic Co-operation and Development. (2010a). *PISA 2009 Results: What Students Know and Can Do — Student Performance in Reading, Mathematics and Science Volume I*. Paris: Author.

Organization for Economic Co-operation and Development. (2010b). *PISA 2009 Results: Overcoming Social Background Equity in Learning Opportunities and Outcomes, Volume II*. Paris: Author.

Organization for Economic Co-operation and Development. (2010c). *PISA 2009 Results: Learning to Learn — Student engagement, strategies and practices, Volume III*. Paris, Author.

Organization for Economic Co-operation and Development. (2010d). PISA 2009 Results: *What Makes a School Successful? Resources, Polices and Practices, Volume IV*. Paris: Author.

Organization for Economic Co-operation and Development. (2010e). *PISA 2009 Results: Learning Trends Changes in Student Performance Since 2000, Volume V*. Paris: Author

Organization for Economic Co-operation and Development. (2010h). *Mathematics Teaching and Learning Strategies in PISA.* Paris: Author.

Scheerens, J., and Bosker, R.J. (1997). The Foundations of Educational Effectiveness. *International Review of Education*, Volume 45, Number 1, 113-120.

Wang, M. C. Haertel, G. D. and Walberg, H. J. (1993).Toward a Knowledge Base for School Learning. *Review of Educational Research*, 63, 3, 249-294.

Willms, J.D. (2003). *Ten Hypotheses about Socioeconomic Gradients and Community Differences in Children's Developmental Outcomes*. Ottawa: Applied Research Branch of Human Resources Development Canada.

Willms, J.D. (2004) *Reading Achievement in Canada and the United States: Findings from the OECD Programme for International Student Assessment*. Ottawa: Human Resources and Skills Development Canada. Retrieved from http://www.pisa.gc.ca/eng/pdf/sst2E.pdf

Willms. J.D. (2006). *Learning Divides: Ten Policy Questions about the Performance and Equity of Schools and Schooling Systems*. Montreal, UNESCO Institute for Statistics.

Wöessmann, L., Lüdemann, E., Schütz, G., & West, M. R. (2007). *School accountability, autonomy, choice, and the level of student achievement: International evidence from PISA 2003.* Paris: Organisation for Economic Co-operation and Development.

TABLE A.1 Regression Coefficients for Populations

Label	Bivariate		Populations		Student demographics		Teacher characteristics		School characteristics		Student reading strategies		Instructional climate		Time allocation and use		Teaching strategies in reading		Assessment	
	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE
Intercept			**504**	**3.67**	**357**	**9.44**	**369**	**10.30**	**344**	**13.90**	**343**	**15.73**	**357**	**18.57**	**356**	**19.92**	**441**	**29.13**	**430**	**34.56**
BCe	-7.32	4.01	-15.97	5.01	-15.41	4.51	-17.19	4.42	-19.64	4.70	-18.38	4.60	-19.80	4.68	-16.74	5.03	-17.51	4.91	-14.42	5.27
ABe	-4.17	3.83	-13.20	4.88	-15.57	4.37	-17.27	4.33	-13.56	4.43	-11.67	4.31	-12.35	4.29	-10.80	4.67	-9.86	4.66	-12.10	5.04
SKe	-26.21	3.36	-33.98	4.55	-31.31	4.15	-31.20	4.13	-22.35	4.75	-21.46	4.60	-23.25	4.63	-21.22	5.31	-21.44	5.25	-17.49	5.83
MBe	-14.78	4.36	-27.34	5.15	-27.18	4.61	-28.17	4.55	-22.03	5.02	-20.88	4.85	-22.07	4.93	-19.66	5.22	-16.89	5.09	-15.66	5.53
ONe*	17.52	3.99	-29.09	6.99	-41.06	5.95	-38.06	5.87	-23.99	6.01	-14.13	5.79	-15.09	6.06	-20.42	5.86	-14.92	6.29	-11.20	7.36
QCe	-19.25	6.66	-28.22	7.37	-26.56	6.85	-28.63	6.75	-35.18	6.70	-35.91	6.45	-35.73	6.31	-33.18	6.29	-29.43	6.10	-29.69	6.25
NBe	-29.47	4.18	-38.54	5.22	-40.83	4.69	-42.73	4.61	-41.65	4.93	-38.58	4.82	-37.41	4.75	-34.04	5.00	-34.25	4.95	-32.39	5.52
NSe	-24.66	3.83	-33.84	4.95	-30.72	4.43	-32.59	4.33	-29.80	4.79	-27.42	4.65	-27.42	4.59	-25.29	4.69	-25.67	4.59	-21.61	5.28
PEe	-32.76	6.26	-42.11	6.88	-32.34	5.70	-33.32	5.39	-32.25	6.17	-32.17	5.91	-33.87	5.99	-30.82	6.00	-31.31	6.23	-28.76	7.17
NLe	-29.83	4.72	-38.87	5.64	-33.99	5.17	-36.06	5.10	-36.16	5.48	-31.97	5.44	-31.97	5.44	-30.51	5.39	-31.06	5.23	-30.98	6.32
YKe	-10.13	14.68	-21.82	13.04	-29.20	12.32	-29.21	12.08	-11.72	12.26	-9.18	11.18	-8.44	11.39	-7.85	10.60	-6.61	10.36	-2.41	13.20
QCf	40.92	6.15	28.57	6.89	40.82	5.87	38.02	5.78	23.99	6.01	14.13	5.79	15.11	6.11	21.57	5.87	14.92	6.29	11.20	7.36
ONf	-15.49	5.03	-24.72	5.91	-13.89	5.60	-14.02	5.56	-14.79	5.83	-19.20	5.66	-16.89	5.67	-16.35	5.77	-16.02	5.92	-12.97	6.87
NBf	-42.81	5.04	-51.78	5.92	-37.82	5.62	-39.18	5.56	-41.80	6.10	-48.92	6.00	-47.70	5.90	-37.58	6.01	-34.80	6.22	-30.16	7.44
MBf	-71.72	9.58	-85.92	10.51	-72.69	10.30	-73.74	10.20	-63.55	10.56	-67.37	9.98	-68.53	9.98	-62.89	9.99	-60.02	9.86	-61.35	11.66
Westf	-9.22	11.76	-21.03	12.71	-13.53	10.74	-13.68	10.62	-4.38	10.50	-9.69	9.65	-5.58	9.54	-2.11	8.72	-2.30	8.63	-34.04	11.41
Eastf	-20.92	10.61	-30.22	11.07	-16.49	9.55	-17.69	9.37	-14.25	9.69	-20.00	9.66	-18.14	9.84	-18.96	9.27	-16.95	9.77	-3.33	15.19

* Coefficient for ONe relative to QCe. All other coefficients relative to ONe.

TABLE A.2 Regression Coefficients for Other Variables

Label	Bivariate		Populations		Student demographics		Teacher characteristics		School characteristics		Student reading strategies		Instructional climate		Time allocation and use		Teaching strategies in reading		Assessment: Full model	
	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE
Intercept			**504**	**3.67**	**357**	**9.44**	**369**	**10.30**	**344**	**13.90**	**343**	**15.73**	**357**	**18.57**	**356**	**19.92**	**441**	**29.13**	**430**	**34.56**
Gender (male=1)	-20.08	2.21			-16.63	2.10	-16.66	2.11	-16.68	2.12	-13.23	2.08	-13.20	2.08	-12.27	2.08	-12.58	2.08	-6.11	2.72
Grade	26.84	2.55			25.36	2.54	25.11	2.52	24.27	2.55	21.92	2.44	21.87	2.43	20.83	2.41	20.62	2.41	15.90	2.93
Not born in Canada	9.63	4.36			0.07	4.39	0.07	4.40	0.51	4.43	4.20	4.08	4.04	4.08	3.99	4.24	4.86	4.20	2.58	5.67
Language match	29.93	3.62			14.17	3.93	14.24	3.93	14.51	3.97	17.58	3.77	17.54	3.76	13.52	3.82	14.21	3.77	7.53	4.42
Books in home	11.43	1.07			18.49	0.98	18.47	0.98	18.29	0.99	11.34	1.04	11.27	1.04	11.02	1.02	10.95	1.02	10.50	1.23
Teacher LA specialization	-10.06	4.02					-7.30	3.57	-6.64	3.55	-6.82	3.43	-6.86	3.44	-7.22	3.36	-6.61	3.43	-10.60	3.88
School enrolment	12.02	2.27							5.70	2.54	4.96	2.49	4.15	2.55	3.80	2.52	3.12	2.54	2.49	2.80
School Governance (Private=1)	35.52	5.93							27.39	5.25	25.01	4.81	23.05	4.95	18.52	4.96	14.02	5.04	16.36	5.61
% Aboriginal enrolment	-13.36	2.30							-7.18	2.21	-6.74	2.09	-5.74	2.12	-4.46	2.11	-4.42	2.04	-3.08	2.40
Community size	3.29	1.32							-1.31	1.29	-1.50	1.28	-1.45	1.26	-1.39	1.25	-1.17	1.20	-0.60	1.33
Reading by decoding	-1.69	0.11									-1.49	0.11	-1.48	0.11	-1.35	0.10	-1.35	0.10	-1.25	0.13
Reading outside of class	2.76	0.12									2.09	0.13	2.09	0.13	2.19	0.13	2.18	0.13	2.12	0.16
Class size	7.85	2.10											1.72	1.90	1.20	1.88	1.04	1.85	0.95	2.12
Grades in class	-4.34	3.34											-0.90	2.81	-0.41	2.83	-0.31	2.75	2.63	3.07
Adjust teaching strategies for special needs	-9.59	2.72											-6.49	2.47	-5.61	2.57	-5.07	2.54	-5.94	2.82
Minutes per week LA	-0.03	0.01													-0.02	0.01	-0.02	0.01	-0.01	0.01
School average absence rate	-7.86	3.13													-3.42	2.74	-2.78	2.71	-1.60	2.98
Student absence rate	-3.86	1.05													-3.12	0.96	-3.08	0.96	-3.29	1.16
Teacher expected homework time	10.71	2.67													4.55	2.14	3.69	2.19	5.67	2.48
Homework counts towards marks	-10.92	2.64													-2.49	2.64	-1.91	2.62	0.28	3.35
Hours per week homework	2.76	0.81													2.92	0.78	2.74	0.79	2.48	1.08
Student being tutored	-24.62	2.79													-30.58	2.77	-30.32	2.75	-26.57	3.37
Teacher use of creative reading material	0.64	0.22															0.11	0.21	0.17	0.25
Teacher assignment of reading outside of class	12.12	3.75															2.47	3.16	-1.49	3.50
Reteach basic skills	-2.65	2.81															-0.42	2.40	-0.59	2.75
Teacher use of media	-3.00	0.41															-1.65	0.36	-1.49	0.38
Student reported use of short test items	-0.98	0.13																	-0.51	0.12
Teacher assessment by short items	-0.76	0.22																	-0.23	0.23
Teacher assessment by long items	0.26	0.25																	-0.02	0.24
Teacher use of assignments for assigning grades	-4.26	1.56																	0.24	1.70
Non-academic criteria for assigning grades	-8.30	1.55																	-2.45	1.54
Student knows what a rubric is	29.38	2.28																	52.77	4.43
Use of rubrics for marking	18.32	2.11																	6.62	1.82
Teacher assessment skill	8.13	3.78																	-1.55	3.46